REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université Mohamed Khider – BISKRA
Faculté des Sciences Exactes, des Sciences de la Nature et de la Vie
Département d'informatique

N° d'ordre : Numéro/M2/2020

Mémoire

Présenté pour obtenir le diplôme de master académique en

Informatique

Parcours : **IA**

Contribution au dépistage intelligent du cancer du sein basé sur la thermographie médicale

---

Par :

HAFIDI MOHAMED MADANI

Soutenu le septembre 2020, devant le jury composé de :

Nom et prénom	Grade	Président
Terrissa Sadek Labib	**Professeur**	**Rapporteur**
Nom et prénom	Grade	Examinateur

Université Mohamed Kheider Biskra
Département d'informatique
Algérie
23 Septembre 2020

Contribution au dépistage intelligent du cancer du sein basé sur la thermographie médicale

Hafidi Mohamed Madani

Mémoire présenté pour l'obtention de diplôme de master académique en Informatique

Résumé

Récemment, l'intelligence artificielle a envahi tous les domaines de la recherche scientifique, pour ce quelle apporte comme solutions intelligentes. Le domaine de la santé n'en fait pas exception.
Le cancer du sein est une maladie à la fois la plus fréquente et la plus mortelle chez la femme à travers le monde et en Algérie. Plusieurs techniques de dépistage de cancer de sein sont disponibles, elles présentent autant d'avantages que d'inconvénients. La mammographie est la technique la plus utilisée pour le diagnostique du cancer du sein. Malgré son efficacité, elle reste inaccessible à toutes les femmes et son coût est relativement élevé. De plus elles est nocive pour la santé des femmes à cause de leurs expositions aux rayonnements X. Une autre technique qui pourrait être une alternative pour le diagnostique précoce du cancer du sein est la thermographie médicale. Cette technique consiste à exploiter la température du sein pour la détection de la tumeur. Contrairement à la mammographie, cette technique moins coûteuse est considérée comme non invasive et non douloureuse. Notre étude consiste à proposer une solution intelligente basée sur la technique de la thermographie médicale pour la détection précoce du cancer de sein chez les femmes.

Mots-clés :Cancer du sein ; diagnostique médical ; Apprentissage automatique ; Réseau de neurones multicouches.

Abstract

Artificial intelligence has recently entered all fields of scientific science, offering intelligent solutions. The area of health is no exception to that. breast cancer is both the most common and deadly illness across the globe and in Algeria, . There are many breast cancer screening approaches available that have as many benefits as drawbacks. The most commonly used method for diagnosing breast cancer is mammography. However, It remains inaccessible to all women , despite its quality, its cost is relatively high. Additionally, because of their sensitivity to X-rays, it is detrimental to women's health. Medical thermography is another approach which may be a choice for the early detection of breast cancer. This method involves using the breast temperature for the tumor to be identified. This less costly procedure, unlike mammography, is known to be non-invasive and painless. Our research consists of proposing an intelligent approach for the early detection of breast cancer based on the technique of medical thermography.

Keywords- Artifical intelligence; Breast cancer; Medical diagnostic; Machine learning; Multilayer neuron networks.

Remerciements

J'exprime ma gratitude et mes remerciements à mon encadrant Professeur Terrissa Sadek Labib pour ses conseils et son soutien pendant le projet, ainsi que pour m'avoir donné l'occasion de réaliser ce merveilleux projet, qui m'a également aidé à faire beaucoup de recherches et à découvrir tant de nouvelles choses.
Je voudrais également remercier ma famille et mes amis qui m'ont beaucoup encouragé à finaliser ce projet dans les délais fixés.

Table des matières

Table des figures

Liste des tableaux

Acronymes

ACR American College of Radiology. 18

CCI Carcinome canalaire infiltrant. 15

CCIS Carcinome canalaire in situ. 15

CLI Carcinome lobulaire infiltrant. 15

DITI Digital Infrared Thermal Imaging. 19

mns minutes. 52

RNMC Réseau de neurones multicouches. 71

SAD Système d'aide à la décision. 2, 22

SADM Système d'aide au diagnostique Médical. 2, 22, 71

Chapitre 1

Introduction Générale

Aujourd'hui, grâce à le développement des ordinateurs et leur capacités d'exécuter des tâches de hautes calculs, plusieurs techniques et algorithmes d'apprentissage ont été implémentés afin d'automatiser des tâches ou d'aider l'être humain à prendre des décisions. L'application de ces techniques est utilisée dans plusieurs domaines.
Le domaine de la santé est l'un des domaines qui s'est bien développé grâce à l'automatisation des tâches par les machines.
Plusieurs systèmes ont été crées pour diagnostiquer des maladies comme le cancer du sein en analysant les images des testes de dépistage pour aider les cliniciens et les médecins à prendre des décisions.
la majorité de ces systèmes utilisent l'apprentissage automatique pour extraire des informations cliniques à partir des images et proposer un diagnostique.
On s'intéresse dans ce projet à la réalisation d'un système intelligent de diagnostique médical du cancer du sein basé sur la thermographie médicale. Les températures du sein étant récupérées par le biais d'un dispositif d'enregistrement sont traitées et appliquées à un réseau de neurone multicouches permettant de localiser une tumeur dans le sein. les résultats obtenus ont été présentés et discutés.
la structure du mémoire est présentée comme suit : Après une introduction générale, un deuxième chapitre consacré à l'anatomie du sein, le cancer du sein et les différentes techniques de dépistage.
Le Chapitre trois introduit le machine learning ainsi que ces techniques. Dans le chapitre quatre, nous présenterons la conception de notre solution et le modèle du classifieur utilisé et dans le chapitre cinq, les résultats obtenus sont discutés. Nous terminons avec une conclusion générale.

Chapitre 2

Cancer du sein

2.1 Introduction

Le cancer du sein chez les femmes représente l'un des grands problèmes de santé à travers le monde.Il est la cause commune de cancer chez les femmes dans les milieux à hauts ressources et à faibles ressources(indépendemment de leur niveau social).Il représente un million parmi les six millions de néoplasmes (tumeurs) diagnostiqué chaque année dans le monde.[12].
Selon les données nationales des registres du cancer en Algérie,environ 44 000 nouveaux cas de cancer ont été enregistrés en 2017, dont 25 000 de ces cas concernent la femme, avec une fréquence très importante du cancer du sein, comparés aux cas enregistrés en 1995 où seulement 300 cas de cancer du sein ont été enregistrés,ce fléau n'a pas cessé de s'accroître dans tous le territoire national.[5]
Les spécialistes, qui jugent inquiétante la prévalence du cancer du sein chez des sujets très jeunes, ont appelé à la mise en place d'un programme national de dépistage précoce.[5]
Dans ce chapitre, on passe en revue le cancer du sein, en définissant le sein et en décrivant ses composants radiologique-anatomique.
On s'étalera aussi sur les différents types de cancers ainsi que sur les techniques de dépistage les plus connues et utilisées dans ce domaine.

2.2 Sein

2.2.1 Définition

Sein (du latin sinus, « courbure, sinuosité, pli ») : Organe pair très développé situé à la partie antérieure du thorax chez la femme, et qui contient la glande mammaire.[1].
Les glandes mammaires sont les glandes des mamelles des femmes, elles ont la particularité de sécréter le lait servant à l'allaitement[14].

2.2.2 Description anatomique du sein

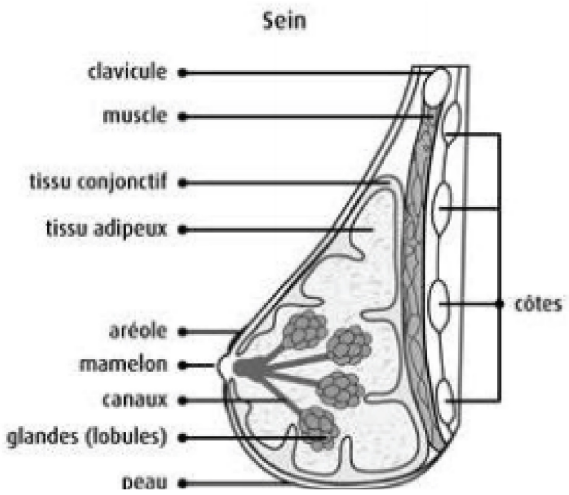

FIGURE 2.1 : Structure anatomique du sein

Le sein est une glande exocrine hormonodépendante qui renferme la glande mammaire. L'élément actif de la glande mammaire est constitué des alvéoles mammaires qui sont regroupées en grappes pour former un lobule [14] (figure2.1).
La paroi des alvéoles est composée de cellules épithéliales sécrétoires et de cellules myoépithéliales contractiles. Chaque alvéole est drainée par un canal dont la paroi est tapissée de cellules épithéliales non sécrétoires et de cellules myoépithéliales [14].
Les canaux de chaque lobule se regroupent en un canal galactophore.
La glande mammaire contient environ une vingtaine de lobes. Chaque lobe regroupe 20 à 40 lobules dont les canaux se déversent vers un canal central,

le canal lactifère. Les lobules sont essentiellement situés à la périphérie de l'organe(par rapport au mamelon)et ils sont plus particulièrement nombreux dans le quadrant supéro-externe. La graisse, présente en quantité plus ou moins importante, et le tissu conjonctif entourent l'ensemble de la glande mammaire [7].

2.2.3 Composition radiologique du sein

De point de vue Radiologique, le sein contient deux composantes :

- Une composante adipeuse radio-transparente.
- Une composante fibroglandulaire radio-opaque.

Ces différentes composantes ont des coefficients d'atténuation aux rayons X presque identiques et donc un contraste peu élevé.[4].

FIGURE 2.2 : Image radiographique d'un sein

Dans une image radiographique d'un sein (figure 2.2), la fibre et la glande apparaissent en blanc alors que la graisse apparaît en noir. donc selon la répartition de ces différentes composantes,les images radiographique des seins seront différentes l'une de l'autre.[4]. On dit qu'un sein est dense lorsque le tissu fibro-grandulaire est plus important que le tissu graisseux [4]. la densité mammaire est mesurée par la proportion des tissus fibro-grandulaire par rapport aux tissus graisseux.
la figure suivante montre les différentes catégories des seins selon le pourcentage de densité mammaire(figure 2.3).

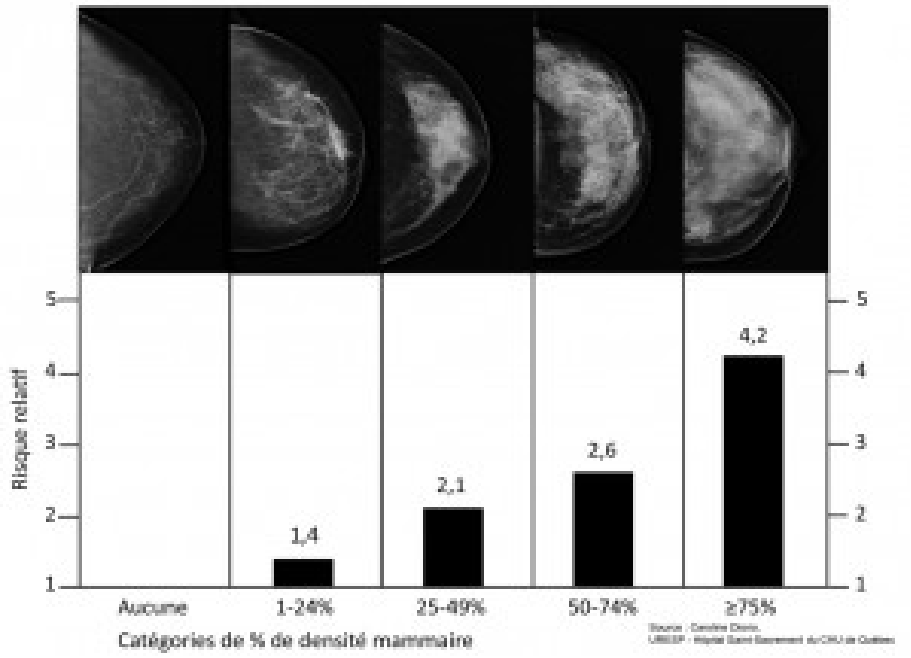

FIGURE 2.3 : Catégorie par pourcentage de densité mammaires des seins

2.3 Cancer du sein

2.3.1 Définition

La notion de ”Cancer du sein” relève d'une nomenclature générique qui fait référence à un ensemble de proliférations néoplasiques de la glande mammaire qui diffèrent tant du point de vue histologique qu'en ce qui concerne leur comportement évolutif[20].
Le terme de “cancer du sein” ne désigne que les tumeurs malignes, potentiellement agressives, du sein tandis que le terme de ”tumeur du sein” désigne à la fois les tumeurs malignes et bénignes. Le terme “carcinome” ou «épithélioma» est parfois utilisé, comme synonyme de “cancer”. En réalité carcinome est un terme spécifique de morphologie microscopique (histologique) qui désigne les plus fréquents des cancers du sein d'origine épithéliale c'est-à-dire dérivant des unités sécrétoires (lobules) ou des canaux lactifères[20].

2.3.2 Facteurs de risque

Âge

D'après les statistiques, 78 % des cancers sont diagnostiqués chez des femmes âgées de plus de 50 ans et moins de 1 % de tous les cancers du sein sont observés chez l'homme[4].

Antécédents personnels, médicaux et familiaux

Une prédisposition génétique. 5 % à 10 % des cancers du sein sont d'origine génétique. Cette prédisposition est liée, le plus souvent, à l'altération génétique des gènes BRCA1 ou BRCA2.Un antécédent personnel de cancer du sein invasif ou de carcinome canalaire in situ, d'hyperplasie lobulaire atypique ou de cancer lobulaire in situ. Un antécédent personnel d'irradiation thoracique médicale à haute dose [4].

Mode de vie (aliments, grossesse, poids...)

La consommation d'alcool et du tabac ont une relation dose-dépendante avec le risque de cancer du sein.Ce risque augmente par absorption de 10g d'alcool par jour[22].Si la femme a commencé à fumer 5 ans avant sa première grossesse [22]. L'obésité augmente le risque de cancer du sein chez la femme ménopausée [22]. D'après certaines études, l'adiposité abdominale élève ce risque [22]. Aussi, des études ont montré que le parabens utilisé dans la fabrication des antitranspirants pourraient simuler l'activité des oestrogènes et de ce fait, favoriser la croissance des cellules cancéreuses mammaires [22].

2.3.3 Symptomatologie

Les symptômes de cancer du sein dépend de son stade de développement, en général, la symptomatologie peut être :

Une grosseur ou une induration au niveau du sein ou de l'aisselle, douloureuse ou non.[22]

Une déformation, une ulcération, une inflammation ou une rétractation de la peau au niveau d'un des seins.[22]

Un écoulement uni-pore sérosanguin au niveau du mamelon, un eczéma ou un érythème de l'aréole ou du mamelon.[22]

2.3.4 Types des cancers du sein

Il existe plusieurs types du cancer de sein (Figure 2.4) :

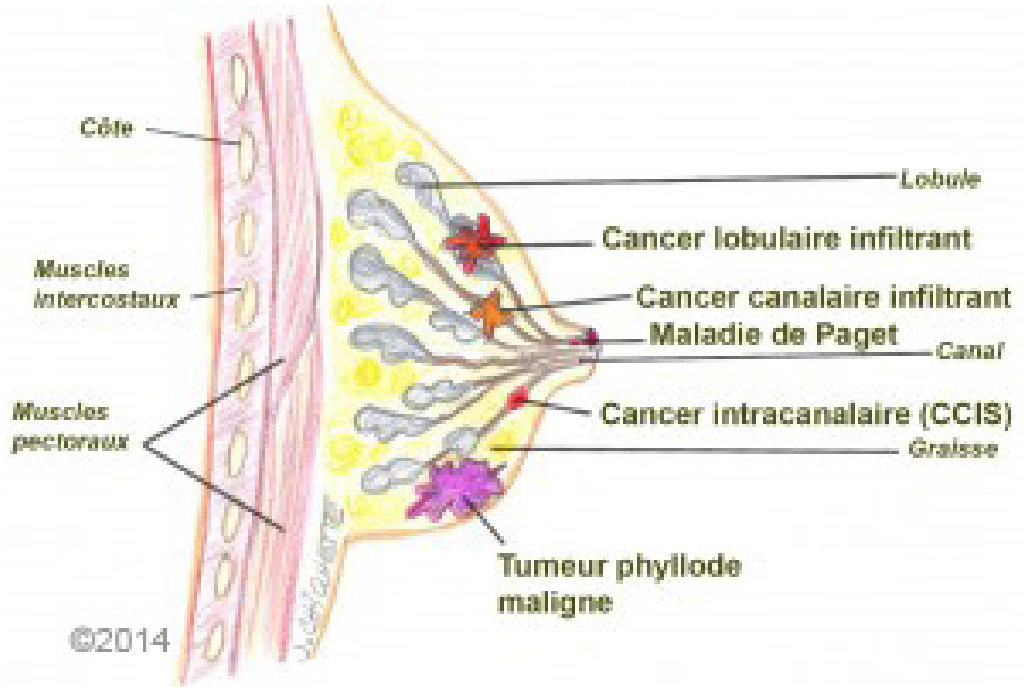

FIGURE 2.4 : Types de cancers du seins

Carcinome canalaire in situ (CCIS)

Renvoie aux cellules anormales repérées dans les canaux galactophores. Ces cellules sont considérées comme non envahissantes puisqu'elles ne se sont pas propagées aux tissus mammaires voisins. Il s'agit d'une forme précoce de cancer qui peut parfois devenir infiltrant et toucher d'autres tissus.

Carcinome canalaire infiltrant (CCI)

Un cancer qui est apparu dans les canaux lactifères et qui s'est propagé aux tissus mammaires environnants.[4]

Carcinome lobulaire infiltrant (CLI)

Un cancer qui a surgi dans les lobules (les groupes de glandes mammaires) et qui a envahi les tissus mammaires voisins.[4]

La tumeur phyllode maligne

Est une tumeur rare du sein, qui prend naissance dans le tissu conjonctif. Il s'agit d'une tumeur mixte, dite fibro-épithéliale, caractérisée par la prolifé-

ration de cellules épithéliales et de cellules du tissu conjonctif, alors que la majorité des cancers du sein affecte les cellules glandulaires.[4]

2.4 Technique de dépistage précoce de cancer du sein

Il existe plusieurs techniques de dépistage précoce de cancer du sein, chaque technique est basé sur une approche différente pour détecter les anomalies au niveau du sein, on distingue parmi ces techniques les deux techniques suivantes :

2.4.1 Mammographie

La mammographie est un examen clinique radiographique du sein, Elle est utilisé pour détecter s'il existe des éventuelles anomalies au niveau des tissus notamment au niveau de la glande mammaire. elle est bilatérale et comporte donc pour chaque sein un cliché de face et un cliché de profil. [4].

Circonstances de prescription

Une mammographie est prescrite soit dans le cadre d'un dépistage du cancer du sein (mammographie de dépistage) soit en présence de symptômes (mammographie de diagnostique) [4].

Appareil de mammographie

L'appareil de mammographie (Figure 2.5) se compose d'un générateur de rayons X d'une faible énergie et d'un système de compression du sein. Il existe deux types d'appareils de mammographie, analogique et numérique.
[7] la mammographie analogique (traditionnelle) imprime l'image sur un film argentique(cliché), par contre la mammographie numérique dessine la valeur numérique de chaque points de l'image pour la former[7].

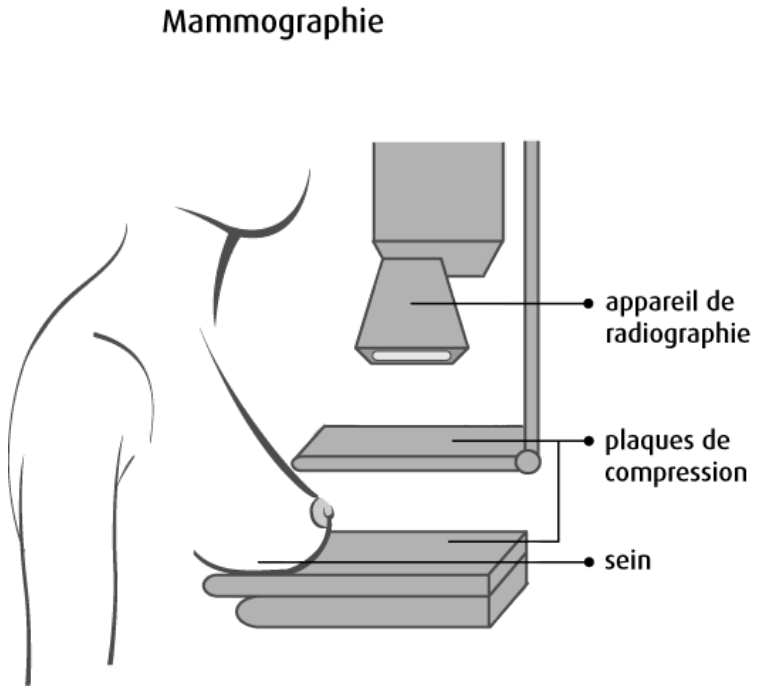

FIGURE 2.5 : Schéma descriptif d'un appareil de mammographie

Description de l'examen

Le radiologue commencera par la palpation manuelle des seins, puis il comprime un premier sein en le plaçant entre les deux plaques de compression. Après la vérification de sa bonne position, le cliché radiographique est pris.le sein est décomprimé et on recommence pour le deuxième sein [7].
L'examen dure de 10 à 20 minutes .
Dans le cas de mammographie diagnostique,l'examen peut prendre plus de temps pour prendre d'autres clichés supplémentaire nécessaire comme grossir certaines régions du sein pour obtenir plus d'informations dans l'image[7].
La veille et le jour de l'examen, aucun déodorant, poudre, crème ou parfum ne doit être appliqué sur les seins ou les aisselles, ces produits pouvant générer des images artéfactuelles pouvant être confondues avec des lésions cancéreuses[4].

Compte rendu de mammographie

Le descriptif et le compte rendu mammographique doivent préciser pour les calcifications le nombre et la taille des foyers, la topographie, la morphologie des calcifications, le nombre de calcifications par foyer, la distribution spatiale, l'évolutivité dans le temps et la présence éventuelle de signes associés,

ainsi que la comparaison avec les mammographies antérieures.[7] La conclusion utilise la classification de l'American College of Radiology (ACR) pour préciser le degré de suspicion de malignité :[7]

- ACR 0 : des investigations complémentaires sont nécessaires.
- ACR 1 : mammographie normale.
- ACR 2 : il existe des anomalies bénignes ne nécessitant ni surveillance ni examen complémentaire.
- ACR 3 : il existe une anomalie probablement bénigne pour laquelle une surveillance à court terme est conseillée.
- ACR 4 : il existe une anomalie indéterminée ou suspecte qui indique une vérification histologique.

2.4.2 Thermographie Médicale

La thermographie médicale des seins est un examen médicale physiologique qui permet de mesurer les différences de température au niveau des seins en une image thermique appelée thermogramme [26]. Le but de l'examen est de détecter des anomalies à partir des températures observés dans un thermogramme. Une émission de températures plus élevées est observée dans les cas de présence d'une masse maligne (cancer) par rapport aux cas normaux à cause de la haute activité métabolique des cellules cancéreuses [26], la figure 2.6 illustre une image thermographique.

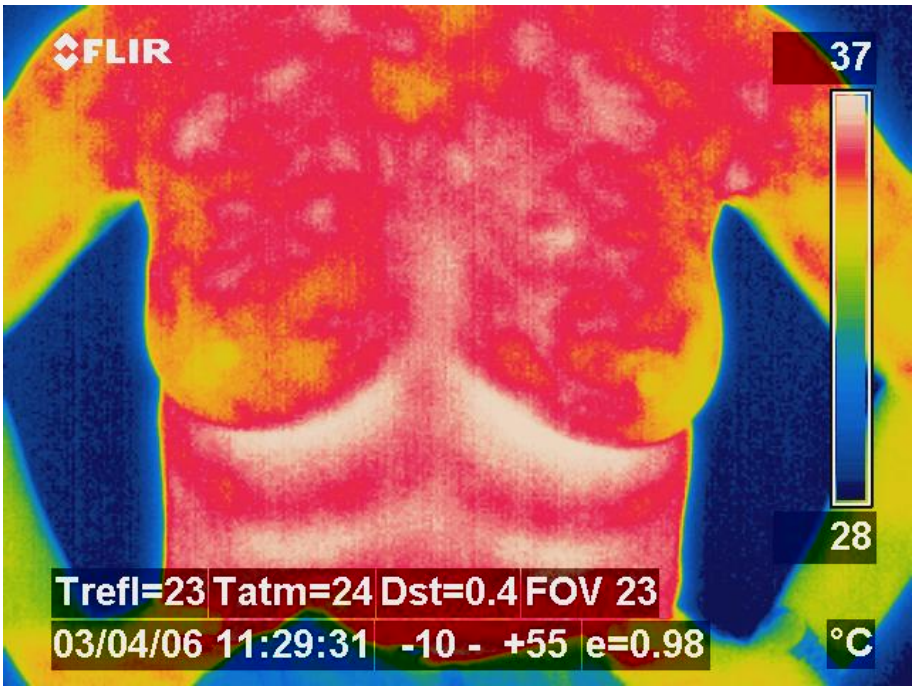

FIGURE 2.6 : Exemple d'un thermogramme (Image thermographique)

Circonstance de prescription

Dans le cadre du dépistage du cancer du sein, on utilise une thermographie spéciale, la Digital Infrared Thermal Imaging. L'examen est non-invasif et non douloureux et ne présente aucun risque de santé pour les patientes, elle permet d'éviter l'exposition aux rayons X [26].

Appareil de thermographie

Une caméra thermique est nécessaire pour prendre une photo de la poitrine et qui sera par la suite traitée par un ordinateur en utilisant les différents algorithmes de traitement d'images thermiques.Généralement les températures dans ces images sont représentées par des couleurs. du bleu (la plus faible température) au rouge (la plus haute température[26].

Description de l'examen

Une caméra thermique est une photo ou image numérique de la poitrine. Les différentes zones de température apparaissent de couleurs variées. Du plus froid (bleu) au plus chaud (rouge). En général, le médecin commence par comparer les images obtenues entre les deux seins. Pour le cancer du sein, ce

sont les vaisseaux sanguins et l'inflammation autour de la tumeur qui sortent en rouge. Mais toute différence de couleur n'est pas forcément un cancer [26]. Cependant, pour ne pas interférer, certaines précautions sont à prendre avant de passer l'examen. Pas de douche chaude avant, pas de rasage ni déodorant, pas de café ou d'alcool. La température de la pièce doit être régulée entre 19°à 23°Celsius. [26].

2.4.3 Conclusion

Dans le chapitre deux, on a définit les concepts théoriques concernant le cancer du sein, on a discuté aussi à propos de ses symptômes et ses facteurs de risques, puis on a décrit ses techniques de dépistage.Cela nous aidera à comprendre beaucoups plus la maladie pour concevoir un système de diagnostique adéquat.
Dans le chapitre suivant, on discutera à propos de l'apprentissage automatique, ses concepts de bases et ses différents types, et on détaillera beaucoups plus sur le principe de fonctionnement de la technique utilisée dans notre projet.

Chapitre 3

Machine Learning (Apprentissage Automatique)

3.1 Introduction

L'intelligence artificielle est l'un des domaines les plus actifs de nos jours en informatique. Elle représente l'avenir de l'informatique classique.
On peut définir l'intelligence Artificielle comme la faculté de reproduire un raisonnement par des moyens informatiques pour des fins décisionnelles[23].
L'intelligence Artificielle comporte plusieurs champs d'applications tels que : la traduction automatique, jeux de réflexions, démonstration des théorèmes mathématiques, diagnostiques de maladies, prédiction, etc., et chacun suit une approche spécifique pour la mettre en oeuvre.
On distingue deux approches principales en Intelligence Artificielle [23] :

- Approches symboliques : tels que les systèmes experts , raisonnement par cas,...etc.
- Approches connexioniste : comme l'apprentissage artificiel (réseaux de neurones artificiel)

L'apprentissage automatique (Machine Learning en Anglais) est un sous domaine de l'intelligence Artificielle connexionniste, il se base sur des approches statistiques et mathématiques pour faire apprendre l'ordinateur à effectuer une tâche à partir des données (base de données) et à travers une phase d'apprentissage.

3.2 Intelligence Artificielle et médecine

3.2.1 Système d'aide à la décision SAD

La prise des décisions dans les systèmes complexes est souvent basée sur les capacités cognitives des être humains.[6].
Dans les systèmes d'aide à la décision, la qualité des prédictions et leur validité est importante. L'aide des convenances du jugement humain et de prise de décision ont été une préoccupation majeure de la science à travers l'histoire. Plusieurs techniques dans le domaine de statistique ont été développées pour faciliter la prise des décisions, ces techniques ont été plus tard renforcées par des méthodes de l'intelligence artificielle tels que les réseaux de neurones artificiels. Le concept des SADs est extrêmement vaste, sa définition dépend selon le sujet traité [6].
On peut alors définir approximativement les SADs comme des systèmes complexes interactifs basés dans une machine qui aide les utilisateurs à prendre une décision.
Les SADs ont plusieurs domaines d'application tels que l'ingénierie, Marketing, Médecine, etc.. Ils s'avèrent d'une importance primordiale dans beaucoups de situations notamment dans le cas où les données sont massives et on arrive parfois à des niveaux de précision et d'optimalité très intéressants, ce qui est très utile pour assister les êtres humains. [6]

3.2.2 Système d'aide au diagnostique Médical SADM

L'utilisation d'outils informatiques et l'Intelligence artificielle pour l'aide au diagnostique médical a débuté depuis un demi-siècle [6], Les SADMs sont des applications importantes pour la reconnaissance et l'analyse des données médicales, visant à aider les médecins pour prendre des décisions de diagnostique[8]. Ils tendent à fournir aux cliniciens des informations utiles après avoir décrit la situation clinique du patient, afin de les aider à améliorer la qualité des soins. Les SADMs représentent un moteur d'induction qui apprend les caractéristiques de décision des maladies et peut ensuite être utilisé pour diagnostiquer des futurs patients avec des états pathologiques différentes. Plusieurs algorithmes ont été développés pour construire des SADMs dans plusieurs spécialités médicales. Nous distinguons deux types de méthodes issues de l'intelligence artificielle : les méthodes symboliques (ou la procédure de diagnostique produite peut être écrite sous forme de règles) et les méthodes connexionnistes (ou la procédure de diagnostique produite est de type «boite noire») [8].

3.3 Machine Learning

3.3.1 Définition

Machine Learning (branche de l'IA) figure 3.1 : signifie l'apprentissage de la machine, appelé en français Apprentissage Artificiel ou Automatique, selon la définition de Tom Mitchell, c'est l'étude des algorithmes qui permettent aux programmes de s'améliorer automatiquement par expérience. De notre

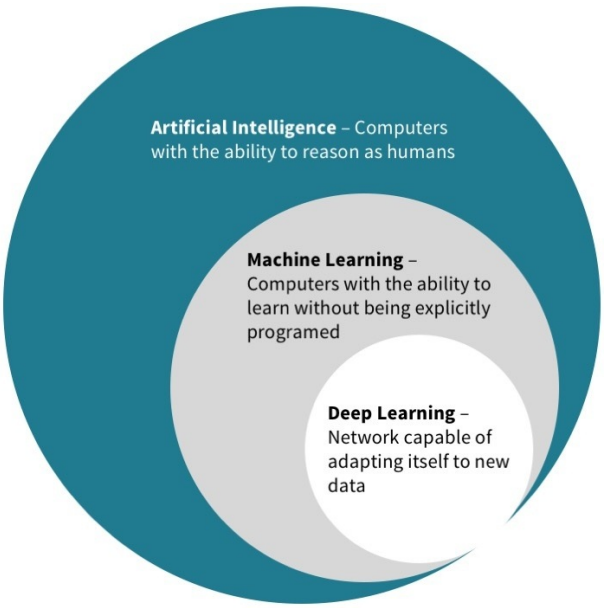

FIGURE 3.1 : Intelligence artificielle vs. Machine learning

point de vue on peut le définir comme l'ensemble de techniques permettant à une machine d'apprendre, à résoudre ou réaliser une tâche sans avoir la programmer explicitement. Cet ensemble de techniques concerne l'analyse, la conception, le développement et l'implémentation de méthodes permettant la machine à suivre un processus systématique pour résoudre un problème, où ça sera difficile ou impossible de le faire par une méthode algorithmique classique. La capacité de faire apprendre une machine à effectuer une tâche nécessite un ensemble de données qui contient pleins d'informations spécifique à cette dernière, les techniques de l'apprentissage automatique permettent d'analyser et traiter toute ces données afin d'extraire les connaissances et les caractéristiques trouvées puis les appliquer et les réutiliser sur de nouvelles données.

3.4 Types d'apprentissage automatique

L'apprentissage automatique a plusieurs approches qui permettent d'extraire les caractéristiques cachées dans les données , on peut classer ces techniques en 3 grandes catégories :

- Apprentissage Supervisé.
- Apprentissage non supervisé.
- Apprentissage semi-supervisé (appelé aussi hybride).

Apprentissage non supervisé

appelés aussi apprentissage sans enseignant, ou apprentissage sans corrélation. Il est utilisé dans les situations où la base d'apprentissage qui contient les différents exemples sont non-étiquetés d'avance. l'apprentissage non supervisé consiste à partitionner les exemples de la base d'apprentissage en catégories en se basant sur un critère de similarité choisi. Il permet la construction automatique des classes sans aucune intervention, mais il nécessite une bonne estimation de nombre de classes[15].

Apprentissage semi-supervisé (hybride)

L'approche hybride consiste à étudier comment les systèmes artificiels (machines) et les systèmes naturels (tels que les humains) apprennent, en utilisant une base d'apprentissage avec des données étiquetés et non étiquetés [15]. ce type d'apprentissage est très performant car il utilise des données non étiquetés pour améliorer le résultat de l'apprentissage supervisé, il permet de comprendre comment le comportement de l'apprentissage change quand on combine les données étiquetés et non étiqueté [15].
Il est souvent utilisé dans les cas où les données étiquetés sont peu et coûteuse à avoir.[15]

Apprentissage supervisé

L'apprentissage supervisé est l'idée d'apprendre à partir d'exemples d'une manière formelle, son objectif est de déterminer comment faire la différence entre les classes des données différentes. Cette approche nécessitent une base d'apprentissage étiqueté, chaque exemple de données a son étiquette qui détermine sa classe d'appartenance, à partir de cette base d'apprentissage on

crée un modèle qui analyse les ressemblances entre les exemples de même classe et les différences entre les exemples de différentes classes dans le but d'avoir la meilleur séparation possible entre les classes de données et pouvoir l'appliquer sur des nouveaux exemples [10].
La figure 3.2 représente une taxonomie des techniques du machine learning.

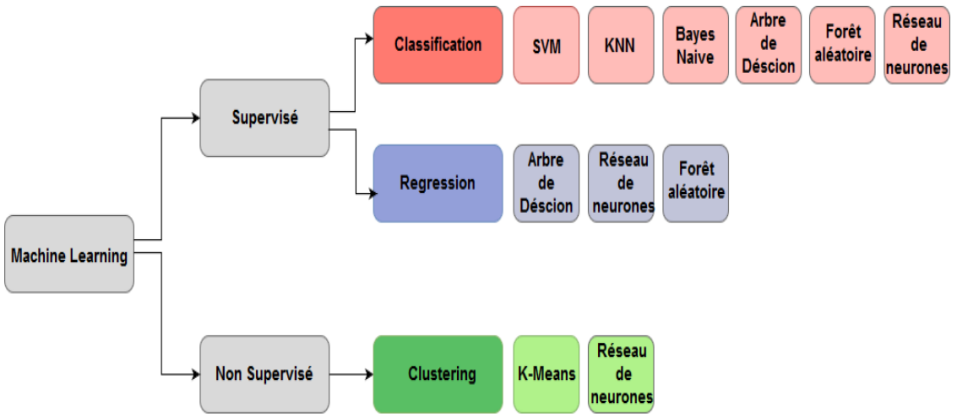

FIGURE 3.2 : Taxonomie des techniques du machine learning

3.5 Modèle

Dans le cas d'apprentissage supervisé, Un modèle est une fonction de décision f qui, prend comme entrée un couple (x, y) tels que $x \in X_i$ où X est l'ensemble des échantillons composés de plusieurs instances (individus, objets,...).
Cet ensemble est l'espace d'exemples. La dimension de cet ensemble est égale aux nombre des attributs $A = (a_1, a_2, a_3, ..., a_d$ qui caractérisent chaque échantillon.
Pour chaque attribut a_i, une valeur $x \in X_i$ est attribuée ou X_i est l'ensemble de toutes les valeurs possibles pour l'attribut a_i. Chaque échantillon est attribué à une valeur y qui représente la cible de la fonction f (étiquette), qui renvoie une prédiction $f(x_i) = \hat{y}_i$. [9].

3.6 Régression

En Apprentissage automatique et pour les problèmes d'apprentissage supervisé, la régression est un modèle où la cible y à prédire est une valeur continue.

3.7 Classification

La classification contrairement à la régression est un modèle qui prédit une valeur cible y discrète.
Il existe plusieurs algorithmes pour effectuer la classification, l'objectif de ces algorithmes est de faire apprendre la machine à classifier les exemples de données correctement. Ainsi, on parle de classifieur généralisé.

3.7.1 Évaluation de classification

La performance d'un modèle de classification est l'un des objectifs de l'apprentissage. Aussi, l'analyse et l'évaluation d'un modèle est essentielle, car elles peuvent être utilisées pour optimiser les valeurs des paramètres du classifieur. Elle permet aussi d'estimer la qualité de généralisation du modèle.
Il existe plusieurs techniques pour évaluer correctement les algorithmes d'apprentissage et la performance prédictive des classifieurs générés.

3.7.2 Technique d'évaluation de la classification

Méthode Hold Out

La méthode Hold Out ou validation croisée simple est une technique qui sert à diviser l'ensemble des échantillons X en deux sous ensembles, le premier sous-échantillon X_a pour la phase d'apprentissage qui représente au moins 60% de l'ensemble des échantillons X et X_t pour l'évaluation qui contient le reste de l'ensemble des échantillons.[13]
On construit en premier le modèle d'apprentissage en l'entraînant sur le sous-échantillon X_a, puis on évalue les performances de modèle entraîné en calculant les différentes mesures d'évaluations sur le sous-échantillon X_t.

Matrice de confusion

La matrice de confusion est une représentation tabulaire des résultats de classification d'un modèle, elle contient les informations sur les classes réelles et les classes prédites des exemples. Cette matrice permet de détecter les erreurs faite par le classifieur et leurs nature, Par conséquent, son rôle est

d'évaluer la capacité prédictive du modèle [13]. La figure suivante présente la matrice de confusion d'un classifieur binaire :

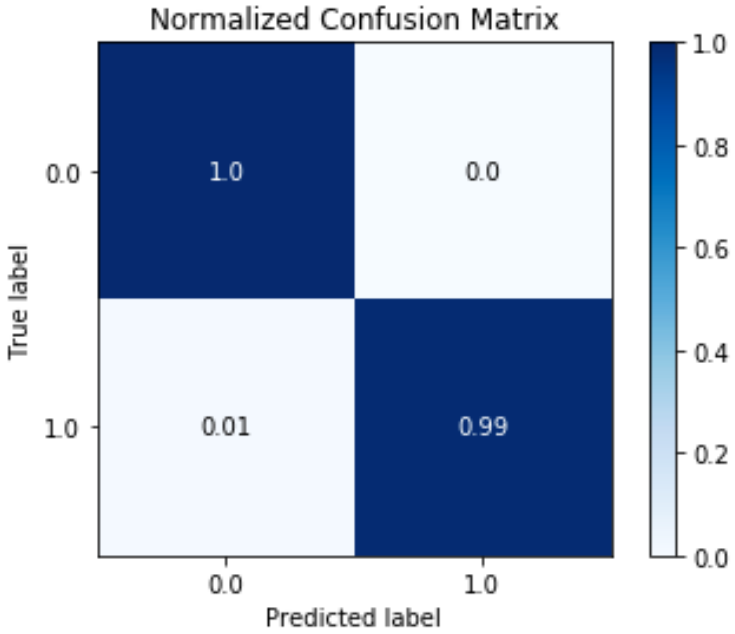

FIGURE 3.3 : Matrice de confusion

à partir de la matrice de confusion, plusieurs mesures d'évaluation peuvent être calculés afin de quantifier les performances de la classification.

Mesures d'évaluation

- **Le taux d'erreur global**
  Le taux d'erreur global E, correspond aux prédictions mal classées, il est calculé à partir de la matrice de confusion, en divisant la somme des coefficients diagonaux qui représente le nombre des prédictions correctes, sur la somme de tous les coefficients de la matrice (le nombre total des exemples), le tout soustrait de l'unité [13] :

$$E = 1 - \frac{Diagonale(M)}{\sum_{i,j} M_{i,j}}$$

- **Précision par rapport à une classe**.
  Cette mesure correspond à la qualité du modèle par rapport à une

classe. On peut la définir comme la proportion d'individus parmi tous ceux pour lesquels le classifieur a prédit cette classe qui appartiennent réellement à celle-ci. Elle se mesure comme le nombre des éléments bien classés par rapport à une classe C, divisé par le nombre total des prédictions attribués à cette classe.[13]
On peut le calculer par la formule suivante :

$$P_C = \frac{M_{C,C}}{\sum_{i=0}^{n} M_{i,C}}$$

- **Sensibilité (Rappel) par rapport à une classe**.
  Elle correspond à la qualité du modèle car elle représente la proportion d'individus que le classifieur a prédit parmi tous ceux qui appartiennent réellement à cette classe, .
  Elle se mesure comme le nombre de des prédictions bien classés par rapport à une classe C, divisé par le nombre total des éléments qui appartiennent réellement à cette classe.
  Elle est donnée par l'équation suivante :

$$R_C = \frac{M_{C,C}}{\sum_{j=0}^{n} M_{C,j}}$$

- **F-mesure par rapport à une classe**.
  On peut résumer les mesures de précision de rappel par rapport à une classe en un seule indicateur, en calculant la moyenne harmonique :

$$F_C = \frac{P_C \times R_C}{P_C + R_C}$$

Pour le cas d'une classification à m classes, on peut résumer ces mesures en calculant la moyenne du rappel, la moyenne de précision et la moyenne des F-mesure comme suit :

$$P_{moyenne} = \frac{\sum_{i=1}^{m} P_{C_i}}{m}$$

$$R_{moyenne} = \frac{\sum_{i=1}^{m} R_{C_i}}{m}$$

$$F_{moyenne} = \frac{\sum_{i=1}^{m} F_{C_i}}{m}$$

3.8 Les Algorithmes d'apprentissage

Dans cette section, nous allons passer en revue les différents algorithmes d'apprentissage dans la classification supervisée, On présentera les principaux algorithmes dans la littérature

3.8.1 KNN le K plus proche voisin(K-nearest neighbour)

KNN est l'un des algorithmes d'apprentissage automatique utilisé pour la classification supervisée. Il fait la classification de chaque nouveau point de données en calculant une métrique de distance (comme la distance Euclidienne,Haming...etc.) entre ce nouveau point de données en entrée et les k plus proches points dans la base de données d'apprentissage (figure 3.4).
L'un des inconvénients de cette approche est le besoin d'un grand espace pour le stockage de l'ensemble complet d'apprentissage, ce qui rend impossible l'analyse des données volumineuse [19].

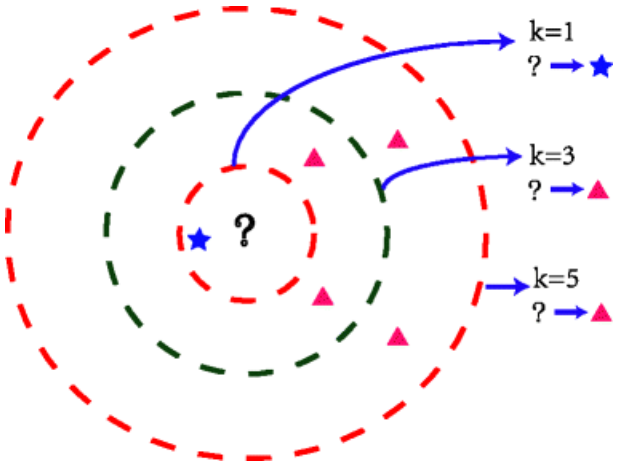

FIGURE 3.4 : K-Nearest neighbour

L'algorithme (comme décrit dans [17]) peut être résume comme suit :

- spécification d'un entier positif k avec un nouvel échantillon.
- sélection des K exemples de données d'apprentissage qui sont les plus proches du nouvel échantillon.
- détermination de la classe la plus commune de ces exemples.
- La classe la plus commune c'est la classe qui va être affectée au nouvel échantillon.

3.8.2 Machine à support vecteurs (SVM)

Les machines à supports vecteurs sont basées sur le concept de plans de décisions qui sépare entre un ensemble d'objets appartenant à une classe.

Ainsi, l'objectif définit les limites de la décision. Cet algorithme prend en charge la classification et la régression et gère les valeurs quantitatives et qualitatives [16].

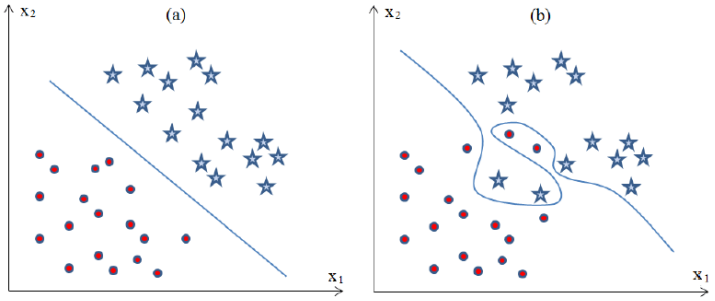

FIGURE 3.5 : Classification par SVM : (a) Problème Linéaire. (b) Problème non Linéaire

3.8.3 Arbre de décision

les arbres de décision sont des techniques qui ont fait un succès en apprentissage dans plusieurs domaines tels que la reconnaissance des caractères, reconnaissance de la parole, diagnostique médical ... grâce à leurs capacités de décomposer un processus complexe en plusieurs sous processus simples facile à prédire et interpréter. L'arbre de décision possède plusieurs algorithmes qui génèrent une structure arborescente où chaque nœud interne désigne un test sur un attribut. Chaque branche représente le résultat du test et chaque nœud feuille contient une étiquette de classe(voir Figure 3.6). Elle vise à partitionner récursivement l'espace d'attribut jusqu'à ce que tous les cas soient complètement partitionnés en sous-ensembles qui ne se chevauchent pas [24]. Parmi les inconvénients de cet algorithme, l'instabilité et le sur-échantillonnage.

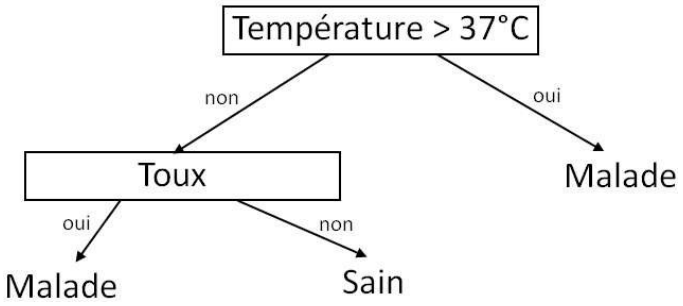

FIGURE 3.6 : Exemple d'arbre de décision

3.8.4 Foret aléatoire

les forets aléatoires représentent un ensemble d'arbres de décision individuelles(voir Figure 3.7), chaque arbre de la foret aléatoire donne une prédiction de classe. La classe qui a plus de vote (qui était prédit le plus) devient la prédiction de la foret aléatoire. L'inconvénient majeure de cette méthode c'est que son résultat est difficile à interpréter et presque non améliorable. L'entraînement de cette méthode est considéré généralement comme lent [27].

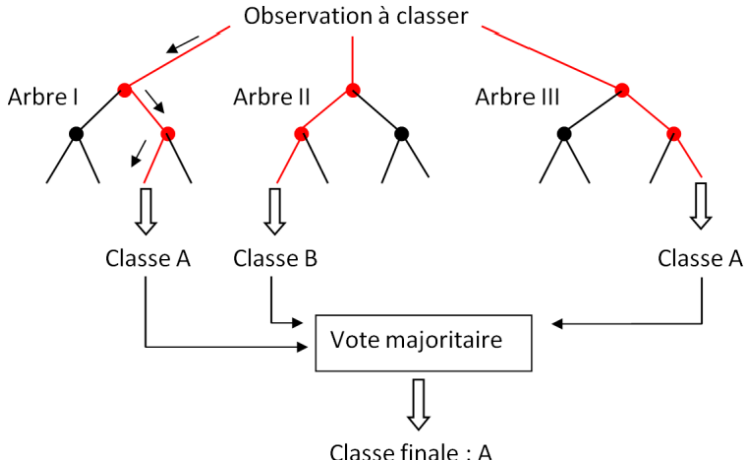

FIGURE 3.7 : Exemple de forêt aléatoire

3.8.5 Réseaux de neurones Artificiels

Les réseaux de neurones sont les outils les plus répandus et utilisés en apprentissage automatique. Ils sont l'une des techniques les plus performantes dans la classification ce qui justifie leur présence presque dans toutes les applications et les domaines qui utilisent l'apprentissage automatique. Leurs structures sont inspirées des neurones biologiques et des circuits du système nerveux du cerveau.
Dans un modèle simplifié du cerveau,les réseaux de neurones consistent d'un large nombre de noeuds de calcul basiques (neurones) qui sont connectés les uns aux autres d'une manière complexe. A travers ces connections, le cerveau est capable de manipuler des calculs de haute complexités.
Un réseau de neurones artificiels peut être décrit comme un graphe orienté tels que chaque noeud représente un neurone et chaque arête représente un lien entre les noeuds. Chaque neurone reçoit en entrée une somme pondérée à partir des sorties de neurones qui le précède.

3.9 Réseaux de neurones artificiels

Pour bien comprendre les réseaux de neurones artificiels, nous aurons besoin de savoir quelques notions de bases des réseaux de neurones biologiques et les adaptation effectuées dans le modèle des neurones artificiels.

3.9.1 Modèle neurophysiologique

Le cerveau se compose d'environ 10^{12} neurones (mille milliards), avec 1000 à 10000 synapses (connexions) par neurone. Ces neurones connectés représente un puissant centre de calcul et de stockage d'information.

Structure d'un neurone biologique

Le neurone est une cellule composée d'un noyau et d'un corps cellulaire. le corps cellulaire se répartie pour former les dendrites,c'est par les dendrites que l'information est transporté de l'extérieur vers le soma, Corps de neurone [25]. L'information traitée continue son chemin à travers un axone pour être transmise à d'autres neurones [25].
La transmission entre deux neurones n'est pas directe, il existe plutôt un petit espace de 10^{-9} entre l'axone de neurone afférant et les dendrites de l'autre neurone. Cette espace défini comme intercellulaire s'appelle un synapse [25].
La figure suivante(figure 3.8) montre la structure d'un neurone biologique :

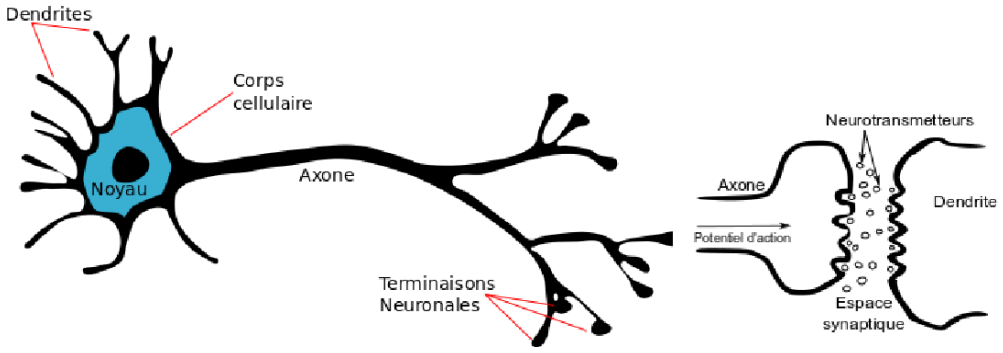

FIGURE 3.8 : Structure d'un neurone biologique

3.9.2 Modèle artificiel

Neurone Artificiel : Perceptron

Le perceptron est un neurone artificiel qui est une simplification aberrante du neurone biologique. Chaque neurone artificiel est un processeur élémentaire. Il prend en entrée un nombre de variables $X = \{x_1, x_2, ..., x_n\}$ appelé couche d'entrée, chaque entrée de neurone artificiel est associée à un poids w (weight en anglais) représentant la valeur de la connexion. Une fonction d'activation f transforme la somme pondérée des variables d'entrées et leurs poids $\sum_{i=1}^{n}(w_i \times x_i)$ à une valeur qui sera ensuite transmise à la couche de sortie pour être comparée à une valeur à seuil, puis fournir une réponse binaire en sortie (0 ou 1) [25]. La figure 3.9 schématise le perceptron.

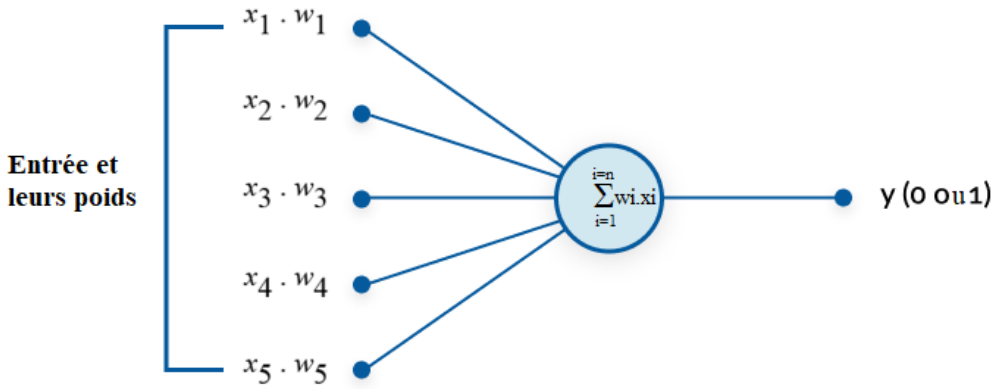

FIGURE 3.9 : Neurone Artificiel perceptron

Perceptron multi-couches

Le perceptron multi-couche est un réseau de neurones artificiel constitué de plusieurs couches de neurones : une couche d'entrée, une couche cachée et une couche de sortie (figure 3.10). Ce modèle est une amélioration du perceptron mono-couche pour remédier aux problèmes non linéaires. L'une des difficultés de ce type de réseaux de neurones et le choix de nombre de neurones dans la couche cachée.

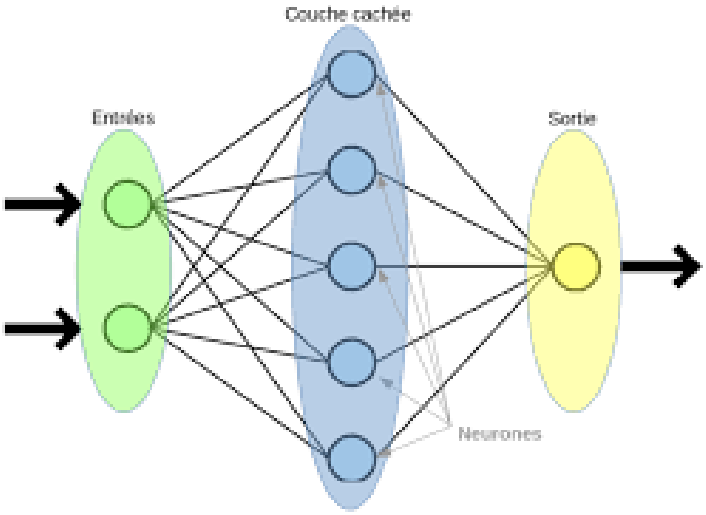

FIGURE 3.10 : Perceptron Multi-couches

Fonction d'activation

Le neurone artificiel calcule la somme pondérée des entrées et leurs poids :

$$\sum_{i=1}^{n}(w_i \times x_i)$$

Cette somme pondérée peut prendre n'importe quelle valeur entre $-\infty$ et $+\infty$, afin que le neurone puisse savoir la valeur seuil pour son activation, une fonction d'activation est utilisée.
La fonction d'activation (ou fonction de transfert) sert à introduire une non-linéarité dans le fonctionnement du neurone contrairement aux neurones biologiques qui ont un état binaire.
Les fonctions d'activation de neurone artificiel ont des valeurs continues ce qui permet d'avoir une infinité des valeurs possibles comprisent dans un intervalle de [-1,1] ou [0,1] [25].
Il existe plusieurs formes de fonctions d'activation, où chacune est utilisée dans un contexte spécifique :

- **Fonction linéaire par morceaux**
  - **Fonction Heaviside** calculée comme suit :

$$f(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0 \end{cases}$$

cette fonction renvoie tout le temps 1 si la valeur en entrée est positif, ou 0 si elle est négative.

- **Fonction ReLU** définie par l'équation suivante :

$$f(x) = \begin{cases} 0 & x < 0 \\ x & x \geq 0 \end{cases}$$

La fonction d'activation ReLU(Rectified Linear Unit) est une fonction linéaire qui est considérée comme la fonction d'activation la plus populaire et utilisée de nos jours, son fonctionnement se résume sur le fait qu'elle remplace toute valeur d'entrée négative par 0, et toute valeur positive par elle même x.[18]

$$ReLU(x) = max(0, x)$$

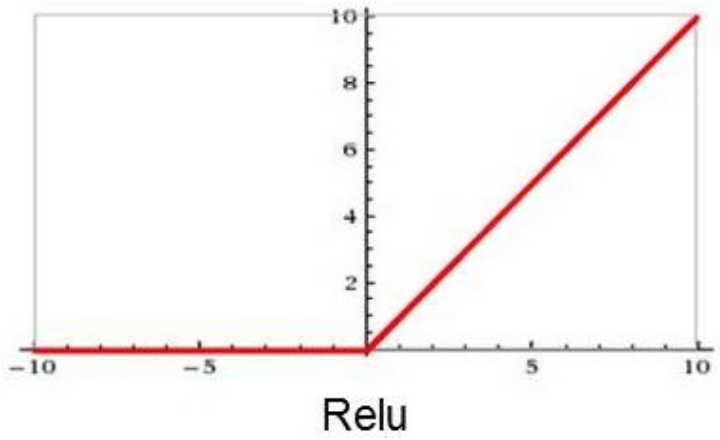

FIGURE 3.11 : Courbe ReLU

- **Fonction sigmoïd**
  - **Fonction sigmoïd** : Définie par l'équation suivante :

$$f(x) = \frac{1}{1 + e^{-x}}$$

Ses valeurs sont dans l'intervalle $[0, 1]$ et sa courbe prend la forme d'un S.(Figure 3.12)
Elle est utilisée dans la couche de sortie pour les problèmes de classification binaire.[18]

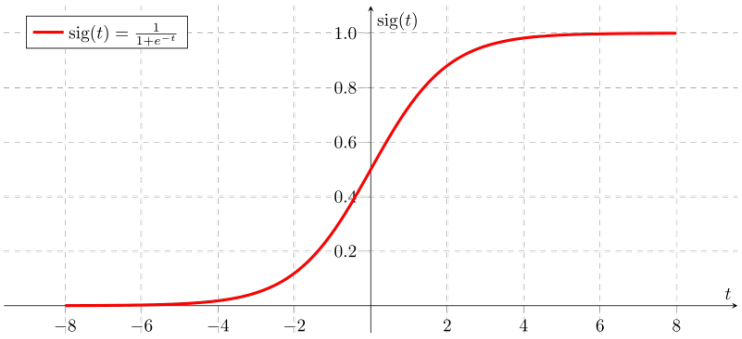

FIGURE 3.12 : Courbe Sigmoïd (sous forme de S)

– **Fonction Tangente-hyperbolique** Cette fonction d'activation transforme toute entrée réelle en une valeur entre [-1,1]. parmi ces inconvénients est le problème de gradient qui devient nul(Vanishing gradient)[18].
Elle est considéré comme une variante de la fonction sigmoïd elle est définie par l'équation suivante :

$$f(x) = tanh(x) = \frac{2}{1 + e^{-2x}} - 1$$

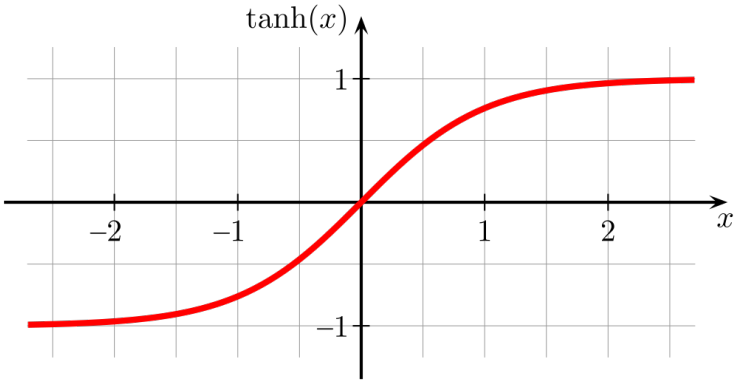

FIGURE 3.13 : Courbe tangente-hyperbolique (sous forme de S)

– **Fonction softmax** ou fonction exponentielle normalisée, elle est utilisée pour représenter une loi catégorielle sur un vecteur $z = (z_1, z_2, .., z_K)$ de K nombres réels en le transformant à un vecteur $\sigma(z)$ à K probabilités de résultats possibles où la somme des K probabilités égale à

1.

$$Softmax(z)_i = \sigma(z)_i = \frac{e^{z_i}}{\sum_{k=1}^{K} e^{z_k}} \quad \text{et} \quad z = (z_1, z_2, .., z_K) \in \mathbb{R}^K$$

Cette fonction est utilisée dans la couche de sortie pour le cas de classification à k classes ($k > 2$) afin de calculer la probabilité à quelle classe une entrée z_i appartient.(la classe avec la plus grande probabilité) [18].

3.10 Conclusion

Dans le chapitre trois, nous avons présenté le machine learning ainsi que ses techniques. Les types d'apprentissage (supervisé, hybride, non supervisé) ont été aussi passés en revue. Nous nous sommes étalé un peu plus sur les réseaux de neurones artificiels car c'est la technique qui sera utilisée dans notre projet. A l'encontre des chapitre un et le chapitre deux qui étaient consacrés à l'aspect théorique de notre projet, le chapitre trois quant à lui détaillera la partie conception.

Chapitre 4

Étude Conceptuelle

4.1 Introduction

Après avoir défini les concepts théoriques concernant le cancer du sein et ses techniques de dépistages, l'apprentissage automatique et le fonctionnement de réseaux de neurones artificiels, nous exposerons dans ce chapitre l'objectif du projet, son architecture et sa conception globale.

4.2 Problématique

Le dépistage précoce du cancer du sein en Algérie présente plusieurs problèmes. En plus de sa complexité, il demeure inaccessible à une tranche importante de femmes pour des raisons sociales, culturelles et économiques.
L'examen de dépistage utilisé en Algérie basé sur la mammographie est recommandé périodiquement chaque deux ans pour les femmes qui dépassent l'âge de 40 ans et chaque année pour celles qui dépassent 50 ans. Deux radiographies par sein sont réalisées, une de face et une en oblique, ce qui permet de comparer les deux côtés de chaque sein.
En cas de détection des anomalies pendant l'examen, des tests complémentaires et une surveillance est obligée pour suivre le développement et le degré de suspicion de malignité.
Nous avons remarqué que la majorité des cancers de sein détectés sont au stade avancé ou final. Cela entraîne généralement des complications importantes voire fatales pour les patientes atteintes de cette maladie. Cette détection tardive est liée à plusieurs facteurs :

- Le manque d'informations chez les femmes notamment dans les régions moins favorisées,

- Les conditions sociales et culturelles qui empêchent les femmes à procéder à la mammographie,
- L'inaccessibilité à la radio mammographique,
- Le coût de la mammographie.

De surplus, le problème de passer une mammographie périodiquement que ça soit dans le cadre de dépistage précoce ou de suivi d'une anomalie probablement bénigne augmente l'exposition aux rayonnements (Rayons X). Les doses cumulées durant toutes les années ne sont pas négligeables et peuvent même devenir cause à développer un cancer du sein. Cette exposition périodique causera aussi une fragilité au niveau de l'ADN, ce qui augmentera le risque génétique de cancer du sein chez les enfants des patientes.

Ces problèmes liés à la fois à la détection tardive et aux effets indésirables de la mammographie, nous ont motivé à chercher des solutions alternatives qui peuvent êtres fiables, moins coûteuses et plus accessibles. La thermographie médicale qui est une technique de dépistage du cancer du sein non-nocive et non douloureuse en fait partie de ces solutions.

4.3 Objectif

L'objectif de cette étude est de concevoir et implémenter une solution alternative de dépistage précoce du cancer de sein chez les femmes. Notre solution est basée sur le principe de la thermographie médicale et les techniques de l'intelligence artificielle.
Il s'agit d'un système de dépistage intelligent qui permet de localiser une tumeur dans le sein. La variation de la température du sein étant enregistrée via un dispositif spécifique représentera l'entrée d'un réseau de neurone artificiel qui est capable de prédire la position de la tumeur dans le sein.

4.4 Architecture générale du système

4.4.1 Composantes du système

La figure 4.1 montre le modèle de dépistage proposé. Il est composé de deux partie : Offline phase et Online phase. La première phase constitue la phase d'apprentissage et la mise à jour des caractéristiques du réseau de neurone artificiel utilisé. Par contre la deuxième phase est celle de de la prédiction de

la position de la tumeur. Ces deux phases utilisent les températures enregistrées depuis le dispositif de mesure.
Les données pré-traitées sont présentées à notre modèle supervisé pour une étape d'apprentissage afin de déterminer les poids du réseau. Plusieurs métriques (Erreur, Exactitude, précision) ont été utilisés pour l'évaluation de la performance de notre modèle.
Une fois la performance de notre modèle est jugée satisfaisante, il peut être utilisé dans la phase online qui permet de prédire la position de la tumeur. Cette phase pourra se servir aussi à la mise à jour de la base de données.

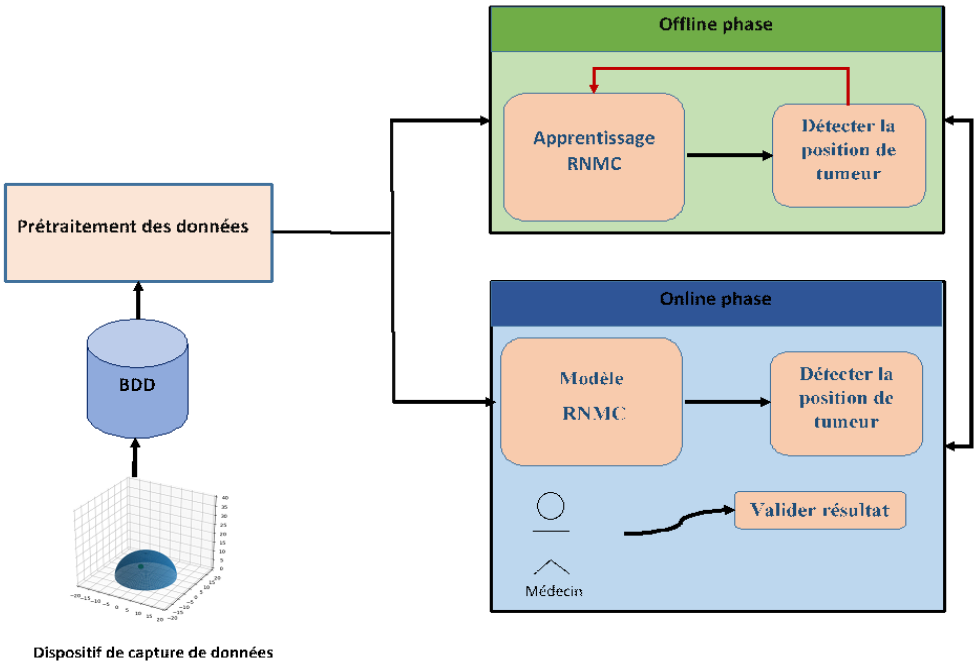

FIGURE 4.1 : Architecture générale du système

4.4.2 Collecte des données

Étant donné que notre approche est guidée par les données, la collection de données est une tâche très importante dans la conception de notre système. Dans notre projet, nous avons exploité un modèle de dispositif de mesure de température de sein où le sein est représenté comme un hémisphère (figure 4.2) composé d'une couche de graisse. La tumeur est représenté comme un cercle avec une position dans le sein ayant ainsi des coordonnées sphériques.

La mesure des températures se fait sur 308 points (capteurs de températures) différents de la surface de sein.

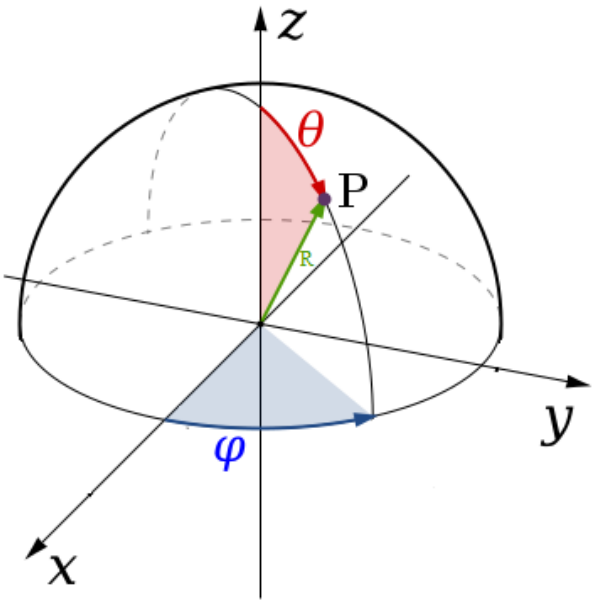

FIGURE 4.2 : Cordonnées sphérique d'une tumeur dans un sein

La base de données collecté est basée sur la mesure des températures de la peau du sein pendant la présence d'une tumeur à une certaine position. La position de la tumeur est identifiée par ses cordonnées sphériques(voir figure4.2) :

- La rotation horizontale α (RotH) qui désigne la longitude (0°et 360°(2π)).
- La rotation verticale θ(RotV) qui désigne la colatitude entre (entre 0°et 90°($\frac{\pi}{2}$)).
- La distance radiale R, qui correspond à la distance (en mm) de l'origine du repère au point P(centre de tumeur).

Le sein est découpé en 4 quadrants, chaque quadrant représente un intervalle de 90 C et est identifié par la valeur de la longitude α. Le premier quadrant Q1 (entre 0°et 90 °), le deuxième quadrant Q2 (entre 90°et 180 °), le troisième quadrant Q3 (entre 180°et 270°) et le quatrième quadrant Q4 (entre 270°et 360°). La collection se fait pour chaque quadrant seule. Pour chaque position, on a 60 exemples de mesures de températures. Chaque quadrant a 560 positions de tumeurs. Donc pour tout le sein , il y aura 2.240 positions de tumeur.

4.4.3 Description de la base de données

Les données issues de notre modèle sont représentées sous forme d'une table contenant plusieurs enregistrements dont le nombre dépend de la fréquence de mesure des températures. Chaque enregistrement est constitué des champs suivants : temps, rotation horizontale, rotation verticale, distance radiale et les valeurs des différents capteurs (figure 4.3).

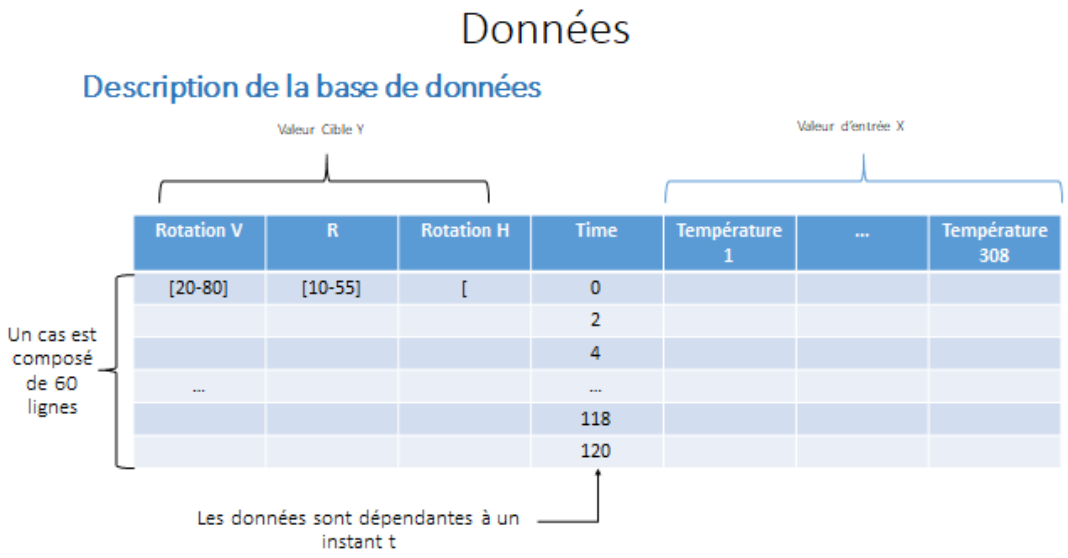

FIGURE 4.3 : Description de la base de données

4.4.4 Pré-traitement des données

L'étape de pré-traitement des données consiste à récupérer l'ensemble des données brutes, les nettoyer et les agréger pour pouvoir les comprendre et les exploiter.
Nous avons normalisé les données pour éviter le problème de différence d'échelle. Cela revient à mettre les données dans un intervalle fixe.

la normalisation min-max des données se fait par la formule suivante :

$$X_{normal} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

tels que :
X_{normal} : représente la valeur de données après sa normalisation
X : représente la valeur de données avant sa normalisation

X_{min} : représente la valeur minimum de données présente dans la colonne des données
X_{max} : représente la valeur maximum de données présente dans la colonne des données

4.4.5 Découpage train/test

L'étape suivante consiste à découper l'ensemble des exemples de la base de données en plusieurs parties, une grande partie des exemples de données est destiné à la phase d'entraînement du modèle (données d'entraînement). La deuxième partie des données(données de test) est utilisé pour calculer la performance et la précision du modèle d'apprentissage afin d'observer et tester les résultats obtenue lors de l'entraînement. Dans la phase de validation, une troisième partie des données peut être utilisée pour valider les résultats du test.

4.4.6 Encodage des valeurs cibles

L'encodage des valeurs de sortie est une étape primordiale pour l'apprentissage. Elle permet de représenter les valeurs cibles sous forme binaire. La représentation binaire se fait sous le principe d'encodage 1 parmi n qui consiste à représenter une valeur x parmi n valeurs possibles sous forme binaire avec un seul 1 entre $(n-1)$ 0. L'avantage de cette technique est que pouvoir passer d'une valeur à une autre, seule deux transitions sont nécessaires (un chiffre passe de 1 à 0 et un autre de 0 à 1).Le tableau 4.1 illustre l'encodage 1 parmi n valeurs cible pour trois bits.

Valeurs cible avant encodage	**1**	**2**	**3**
Valeurs cible après encodage	1	0	0
	0	1	0
	0	0	1

TABLE 4.1 : Encodage valeurs cible

4.4.7 Modèle de prédiction

Le modèle de prédiction de la position de la tumeur choisi pour notre projet est celui des réseaux de neurones artificiels multi-couches (RNMC). C'est un

choix qui est basé à la fois sur l'efficacité de ce type de réseaux dans la prédiction et au fait qu'il n y a pas de travaux antérieurs sur cette problématique.

Architecture du RNMC

Le réseau de neurones multi-couches est un réseau de neurones composé de plusieurs couches successives. Une couche d'entrée, deux couches cachées et une couche de sortie. Le réseau de neurones est un réseau à propagation avant où les neurones sont interconnectés via les poids. La sortie calculée représente la rotation horizontale, la rotation verticale et la distance radiale.

Fonction d'activation

On note $x_1, x_2, ..., x_n \in X$ où X est l'ensemble des exemples et y est la cible de réseau. Un RNMC applique une transformation des variables d'entrée par une fonction d'activation :

$$y = f(x_1, x_2, ..., x_n, w)$$

tels que w est le vecteur qui contient chacun des poids w_{ijk} estimés par le modèle.le i désigne le numéro d'entrée,le j désigne le numéro de neurone,le k désigne le numéro de couche.
w_{ijk} désigne l'i ème entrée de j ème neurone dans l'k ème couche.
Dans le cas d'une classification binaire , le neurone de sortie est muni également de la fonction sigmoïde tandis que dans le cas d'une discrimination à m classes (Y est qualitative),le neurone de sortie utilise une fonction d'activation softmax à valeurs dans R^m et de somme unit. Ces m valeurs sont assimilables à des probabilités d'appartenance à une classe [11].

Apprentissage du RNMC

L'opération d'apprentissage consiste à trouver les valeurs des paramètres w adaptés à notre problème et qui permettent de prédire la meilleur position de la tumeur. Ils sont déterminés par la minimisation de la fonction perte quadratique ou de celle d'une fonction d'entropie en classification [11]. L'équation suivante donne l'erreur quadratique de l'apprentissage.

$$\sum_{i=1}^{n} [y_i - f(x; w)]^2$$

4.5 Detection de la postion de la tumeur

Le système proposé est constitué de trois étages. Il est capable de déterminer la position de la tumeur dans le sein en estimant la rotation horizontale, la rotation verticale et la distance radiale. Ce système se présente sous forme de plusieurs RNMC en cascade (figure 4.4). Après avoir vérifier la configuration de notre RNMC, on commence par prédire la longitude α à partir des mesures de températures, Ensuite on prédit la colatitude θ à partir des mesures de températures et la longitude α, et on finit par prédire la distance radiale R à partir des mesures de températures, la longitude α et la colatitude θ.

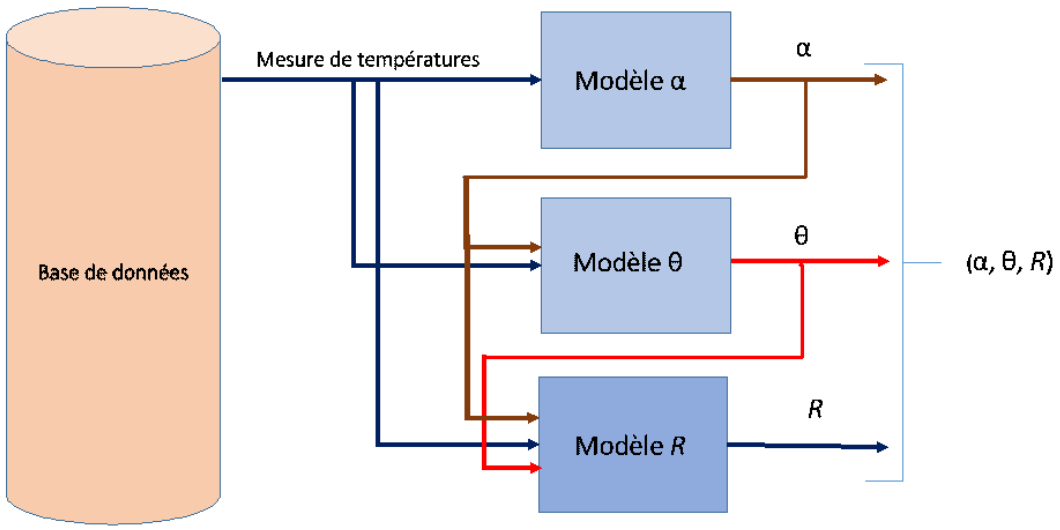

FIGURE 4.4 : Schéma synoptique

4.6 Conclusion

Dans ce chapitre, nous avons présenté la conception de notre système de prédiction de la position de la tumeur. les différents schémas conceptuels ont été

détaillés. Nous avons présenter aussi notre prédicteur neuronal et expliquer le prétraitement des données utilisées. Cette conception sera implémentée dans le chapitre 5.

Chapitre 5

Implémentation

5.1 Introduction

Dans ce chapitre, Nous allons implémenté notre solution intelligente de localisation de la tumeur dans le sein. Cette implémentation consiste à :

- Constituer la base de données à partir du dispositif de mesure de température.
- Prétraiter les données pour améliorer leur qualité et l'adapter au traitement.
- Configurer et tester le réseau de neurones multicouches utilisé dans la prédiction.

5.2 Environnement d'exécution

La simulation étant en version standalone, l'implémentation est réalisé sur un ordinateur ayant les caractéristiques suivantes :

Caractéristique de la machine	Processeur	Intel i7-4510U @ 2.00GHz,2601MHz
	Mémoire Ram	6 GO
	SE	Microsoft Windows 10 Professionnel

Table 5.1 : Caractéristique de la machine utilisée

5.3 Outils et langages de développement

L'implémentation de notre projet nécessite un langage de développement et un éditeur de texte destiné à la programmation. Ce choix a été fait sur le langage Python et l'IDE Pycharm très adaptés à notre problématique.

5.3.1 Python

Python (figure 5.1) est un langage de programmation interprété multi-paradigme qui prend en charge la programmation orienté objet, impérative structurelle et fonctionnelle. Il est considéré comme un langage de haut niveau grâce à ses fonctionnalités avancées tels que la gestion automatique de la mémoire (garbage collecting) [2].

FIGURE 5.1 : Logo Python

En plus de sa simplicité et facilité d'utilisation, le langage Python fonctionne sur la plupart des plates-formes informatiques, des smartphones aux ordinateurs centraux, de Windows à Unix avec notamment GNU/Linux en passant par macOS, ou encore Android, iOS.
Son avantage majeur est sa modularité,la définition de langage est très compacte autour de son noyau, il existe des nombreuses librairies et modules qui ont été développées et qui optimisent la productivité des programmeurs et facilitent la tâche de programmation.
Afin de créer un modèle d'apprentissage , plusieurs librairies peuvent être utilisé , on note les librairies suivantes :

Numpy

NumPy (diminutif de Numerical Python) est une librairie destiné aux langage Python qui permet de stocker et effectuer des opérations sur les données.

D'une certaine manière, les tableaux Numpy sont comme les listes en Python, mais Numpy permet de rendre les opérations beaucoup plus efficaces, surtout sur les tableaux de large taille. Les tableaux Numpy sont au coeur de presque tout l'écosystème de data science en Python [21].(Figure5.2)

FIGURE 5.2 : Logo Numpy

Matplotlib

Matplotlib est une bibliothèque du langage de programmation Python destinée à tracer et visualiser des données sous formes de graphiques. Elle permet l'exportation des graphiques sous plusieurs formes (PNG,JPEG,PDF...) et elle est doté d'une 'User Graphical Interface' qui permet de zoomer et explorer les graphiques facilement.(Figure5.3)

FIGURE 5.3 : Logo Matplotlib

Pandas

Pandas est une bibliothèque écrite pour le langage de programmation Python permettant la manipulation et l'analyse des données. Elle propose en particulier des structures de données et des opérations de manipulation de tableaux numériques et de séries temporelles.
Les principales structures de données sont les séries (pour stocker des données selon une dimension - grandeur en fonction d'un index), les DataFrames (pour

stocker des données selon 2 dimensions - lignes et colonnes), les Panels (pour représenter des données selon 3 dimensions, les Panels4D ou les DataFrames avec des index hiérarchiques aussi nommés MultiIndex (pour représenter des données selon plus de 3 dimensions - hypercube).[21](Figure5.4)

FIGURE 5.4 : Logo Pandas

TensorFlow

TensorFlow est une plate-forme Open Source révolutionnaire de bout en bout dédiée au machine learning. Elle propose un écosystème complet et flexible d'outils, de bibliothèques et de ressources communautaires permettant aux chercheurs d'avancer dans le domaine du machine learning, et aux développeurs de créer et de déployer facilement des applications qui exploitent cette technologie. Elle possède une multitude d'API permettant la construction et l'entraînement des modèles de machine learning, l'une des plus utilisée est L'API Keras (Figure5.5)[21].

FIGURE 5.5 : Logo TensorFlow

Keras

Keras est une bibliothèque Open Source écrite en Python et s'exécute sur TensorFlow. Elle permet d'expérimenter rapidement et facilement les modèles

de réseaux de neurones artifciels. L'avantage de Keras est sa syntaxe assez simple et la richesse des ses fonctions notamment celles dédiées aux modèles de réseaux de neurones (Figure5.6) [21].

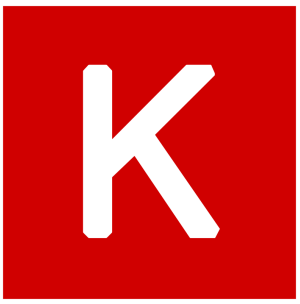

FIGURE 5.6 : Logo Keras

5.3.2 Pycharm

FIGURE 5.7 : Logo Pycharm

PyCharm est un environnement de développement intégré (IDE) développé par JetBrains , utilisé pour programmer en Python.Il permet l'analyse de code et contient un débogueur graphique. Il permet également la gestion des tests unitaires et l'intégration de logiciel de gestion de versions (figure 5.7) [3].

5.4 Réalisation et Implementation

Dans cette partie, on utilisera l'IDE Pycharm et le langage Python afin de réaliser le système conçu dans le chapitre précèdent. ‘

5.4.1 Types et Nature des données

La base de données utilisée est celle qu'on a décrit dans le chapitre précédent. Elle se compose de plusieurs tableaux à deux dimensions. Les données sont collectées à l'aide d'un logiciel de simulation multiphysique dont le type est numérique. l'ensemble des tables est stockée dans un fichier de type .csv (voir Figure5.8).
Une table est constituée des informations concernant la fréquence, la position de la tumeur et les températures enregistrées. Les différents enregistrement sont enregistrés dans un horizon de 120mns avec une fréquence de 2mns. La figure suivante illustre une table de données parmi les fichiers .csv qui composent la base de données.

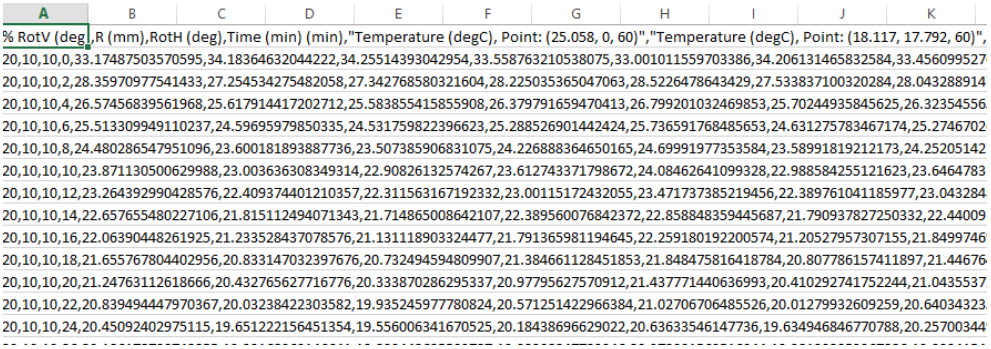

FIGURE 5.8 : Exemple d'un tableau des données

5.4.2 Pré-traitement

Réorganisation et nettoyage de données

Afin de pouvoir amélioré et réorganiser les données, un script python a été écrit pour mettre tout les tableaux dans un Dataframe en supprimant toute les lignes avec les colonnes vides (NaN) et les colonnes non nécessaire pour notre projet. La nouvelle table obtenue est montrée en dessous dans la figure 5.9. Les données ont été enregistrés sous forme de fichier (.npy) afin de minimiser leurs temps de chargement sur python et réduire leurs taille.

Rotation V	R	Rotation H	Time	Température 1	...	Température 308
[20-80]	[10-55]	[10-350]	0			
			2			
			4			
...			...			
			118			
			120			

FIGURE 5.9 : Organisation des données dans le DataFrame

Normalisation des données

Afin d'ajuster les données à une échelle théorique facilitant ainsi son interprétation, nous avons effectué une normalisation des données en utilisant la fonction MinMaxScaler. Les données après la normalisation prennent des valeurs dans l'intervalle [0,1] (voir Figure5.10).

5.4.3 Encodage des valeurs cibles

L'encodage des valeurs cible en binaire dépend de nombre de classes présentes dans la base de données. Dans l'implémentation de notre projet, on a trois valeurs cibles, chacune est prédite seule avec le modèle RNMC. L'encodage des valeurs cibles est fait par la fonction Label Binarizer qui permet de réaliser l'encodage de 1 parmi n expliqué dans le chapitre précédent.
On présentera maintenant les différents changements fait sur la couche de sortie après le résultat de l'encodage.

Encodage des valeurs cible pour le modèle de classification de la longitude α

Après l'analyse de la base de données, précisément la colonne de la longitude (appelés RotH dans notre base de données). nous réalisons qu'on a 32 états de positions pour la colonne RotH (α). Le tableau suivant montre les valeurs possibles de α pour chaque quadrant.

Valeurs possible de α	Q1	10	20	30	40	50	60	70	80
	Q2	100	110	120	130	140	150	160	170
	Q3	190	200	210	220	230	240	250	260
	Q3	280	290	300	310	320	330	340	350

TABLE 5.2 : Valeurs possible de la longitude par quadrant

L'encodage 1 parmi n alors encode la colonne RotH en 32 bits.
Le nombre de neurones de la couche de sortie alors est égale au nombre des états possible = 32.

Encodage des valeurs cible pour le modèle de classification de la colatitude θ

La même analyse sur la colatitude (appelés RotV dans notre base de données). On réalise qu'on a 7 états de positions pour la colonne RotV (θ). Le tableau suivant montre les valeurs possibles de θ pour tout les quadrants.

Valeur de θ pour tout les quadrants	Q1 Q2 Q3 Q4	20 30 40 50 60 70 80

TABLE 5.3 : Valeurs possible de la colatitude par quadrant

L'encodage 1 parmi n alors encode la colonne RotV en 7 bits.
Le nombre de neurones de la couche de sortie alors est égale au nombre des états possible = 7.

Encodage des valeurs cible pour le modèle de classification de la distance Radiale R

En ce qui concerne la distance radiale (appelés Rayon dans notre base de données), elle représente quant a elle 10 états de positions pour la colonne de rayon (R). Le tableau suivant montre les valeurs possibles de R pour tout les quadrants.

Valeurs de R pour tout les quadrants	10 15 20 25 30 35 40 45 50 55

TABLE 5.4 : Valeurs possible de la distance radiale R par quadrant

L'encodage 1 parmi n alors encode la colonne R en 10 bits.
Le nombre de neurones de la couche de sortie alors est égale au nombre des états possible = 10.

5.4.4 Implementation des modèles neuronaux

Afin d'implémenter les modèles RNMC, nous avons utilisé les fonctions prédefinies **Keras.Sequential** et **Keras.Layers** de la bibliothèque **keras**. Une configuration de plusieurs autres paramètres est nécessaire, elle se résume sur le choix **des fonctions d'activations** des différentes couches, **Une fonction d'optimisation** et **une fonction de perte (erreur)**.
Nous obtenons par conséquent un réseau de neurone composé de trois sous réseaux avec des couches différentes de sortie.

5.5 Résultats et discussions

Les résultats obtenus lors de notre implémentation seront analysés et discutés dans cette section. On utilisera les différentes techniques d'évaluations de classification : la matrice de confusion, la mesure d'erreur globale, et le taux d'exactitude. Deux configurations des hyperparamètres sont retenues pour la prédiction de la position de la tumeur. La différence de ces configurations est liée principalement au taux d'apprentissage et le nombre d'epochs.

5.5.1 Longitude α, Colatitude θ et distance radiale R

Longitude α

Pour la longitude α, la première et la deuxième configuration sont données respectivement dans les tableaux 5.5 et 5.7. Les résultats obtenus (précision, rappel, F1-mesure, et support) sont représentés dans le tableau 5.6 pour la première configuration et le tableau 5.8 pour la deuxième configuration.

Première configuration

Hyper-paramètres	**Valeurs**
Nombre de couche cachée	2
Nombre de neurone pour chaque couche cachée	617
Nombre de neurone pour la couche de sortie	32
Fonction d'activation couches cachées	ReLU
Fonction d'activation couche sortie	Softmax
Algorithme d'optimisation	Adam
Taux d'apprentissage (Learning rate)	0.1
Taille du lot (batch size)	60
Fonction d'erreur	Entropie-catégorielle
Nombre d'itérations (epochs)	350

TABLE 5.5 : Configuration de réseau implémenté pour la classification de α

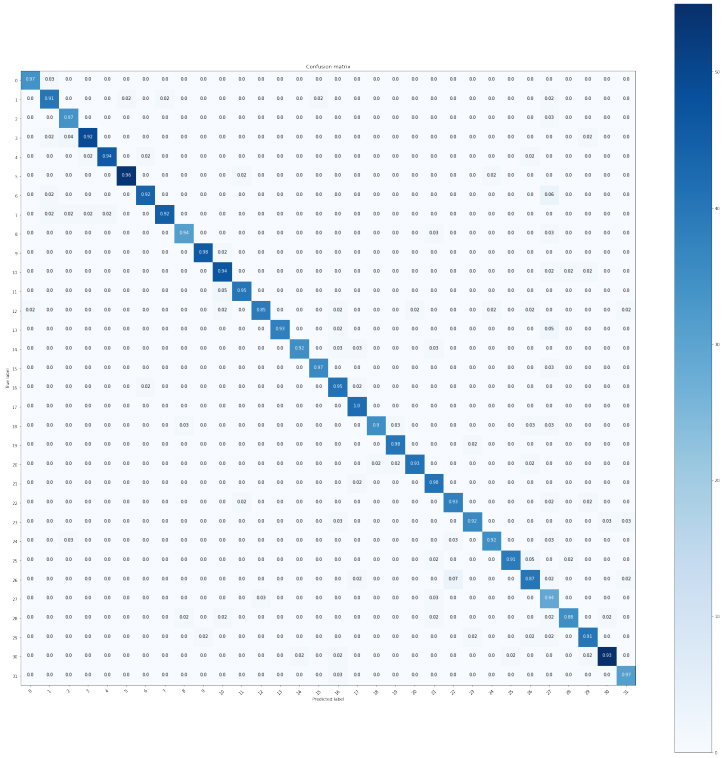

FIGURE 5.10 : Matrice de confusion de la première configuration pour α

Mesures / Classes	**Précision**	**Rappel**	**F1-mesure**	**Support**
10	0.78	0.89	0.83	35
20	0.77	0.84	0.80	44
30	0.83	0.86	0.85	35
40	0.98	0.85	0.91	53
50	0.88	0.86	0.87	49
60	0.98	0.84	0.90	56
70	0.98	0.88	0.92	48
80	0.95	0.82	0.88	49
100	0.96	0.82	0.89	33
110	1.00	0.79	0.88	47
120	0.93	0.86	0.90	50
130	1.00	0.69	0.82	42
140	1.00	0.83	0.90	46
150	0.97	0.78	0.86	41
160	0.94	0.84	0.89	38
170	0.89	0.86	0.88	37
190	0.89	0.95	0.92	43
200	0.88	0.83	0.85	42
210	0.92	0.85	0.88	39
220	0.92	0.85	0.89	41
230	0.83	0.80	0.81	44
240	0.94	0.81	0.87	42
250	0.84	0.78	0.81	41
260	0.94	0.82	0.88	39
280	0.88	0.74	0.80	38
290	0.94	0.79	0.86	43
300	0.80	0.87	0.83	46
310	0.62	0.90	0.74	31
320	0.70	0.78	0.74	40
330	0.55	0.93	0.69	44
340	0.59	0.92	0.72	59
350	0.52	0.84	0.64	32

TABLE 5.6 : Rapport des mesures d'évaluation de la première configuration pour la classification de α

Deuxième configuration

Hyperparamètres	**Valeurs**
Nombre de couche cachée	2
Nombre de neurone pour chaque couche cachée	617
Nombre de neurone pour la couche de sortie	32
Fonction d'activation couches cachées	ReLU
Fonction d'activation couche sortie	Softmax
Algorithme d'optimisation	Adam
Taux d'apprentissage (Learning rate)	ReduceLROnPlateau(0.1 jusqu'à 0.00001)
Taille du lot(batch size)	60
Fonction d'erreur	Entropie-catégorielle
Nombre d'itérations (epochs)	350

TABLE 5.7 : Deuxième configuration de réseau implémenté pour la classification de α

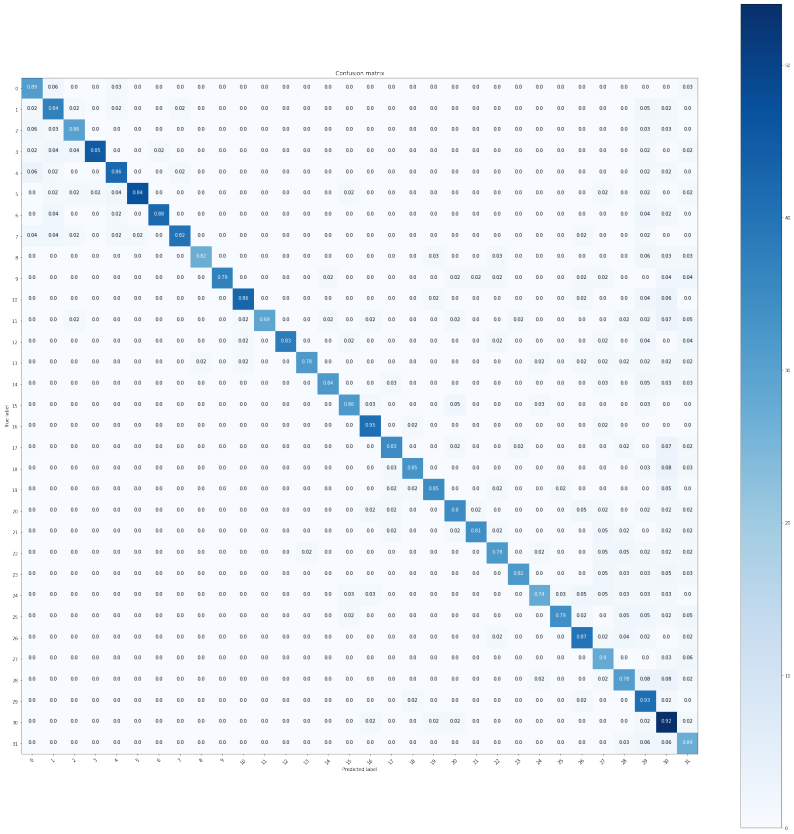

FIGURE 5.11 : Matrice de confusion de la deuxième configuration pour α

Mesures / Classes	**Précision**	**Rappel**	**F1-mesure**	**Support**
10	0.97	0.89	0.97	35
20	0.91	0.84	0.91	44
30	0.89	0.86	0.93	35
40	0.96	0.85	0.94	53
50	0.98	0.86	0.96	49
60	0.98	0.84	0.97	56
70	0.96	0.88	0.94	48
80	0.98	0.82	0.95	49
100	0.94	0.82	0.94	33
110	0.98	0.79	0.98	47
120	0.90	0.86	0.92	50
130	0.95	0.69	0.95	42
140	0.97	0.83	0.91	46
150	1.00	0.78	0.96	41
160	0.97	0.84	0.95	38
170	0.97	0.86	0.97	37
190	0.87	0.95	0.91	43
200	0.91	0.83	0.95	42
210	0.97	0.85	0.93	39
220	0.95	0.85	0.96	41
230	0.98	0.80	0.95	44
240	0.89	0.81	0.93	42
250	0.90	0.78	0.92	41
260	0.95	0.82	0.94	39
280	0.95	0.74	0.93	38
290	0.97	0.79	0.94	43
300	0.85	0.87	0.86	46
310	0.64	0.90	0.76	31
320	0.95	0.78	0.91	40
330	0.91	0.93	0.91	44
340	0.96	0.92	0.95	59
350	0.91	0.84	0.94	32

TABLE 5.8 : Rapport des mesures d'évaluation de la deuxième configuration pour la classification de α

Colatitude θ

La même démarche que celle de la longitude a été faite pour la colatitude θ. les tableaux 5.9 et 5.11 pour les configurations et les tableaux 5.10 et 5.12 pour les résultats obtenus.

Première configuration

Hyperparamètres	**Valeurs**
Nombre de couche cachée	2
Nombre de neurone pour chaque couche cachée	617
Nombre de neurone pour la couche de sortie	7
Fonction d'activation couches cachées	ReLU
Fonction d'activation couche sortie	Softmax
Algorithme d'optimisation	Adam
Taux d'apprentissage (Learning rate)	0.1
Taille du lot (batch size)	60
Fonction d'erreur	Entropie-catégorielle
Nombre d'itérations (epochs)	350

TABLE 5.9 : première configuration de réseau implémenté pour la classification de θ

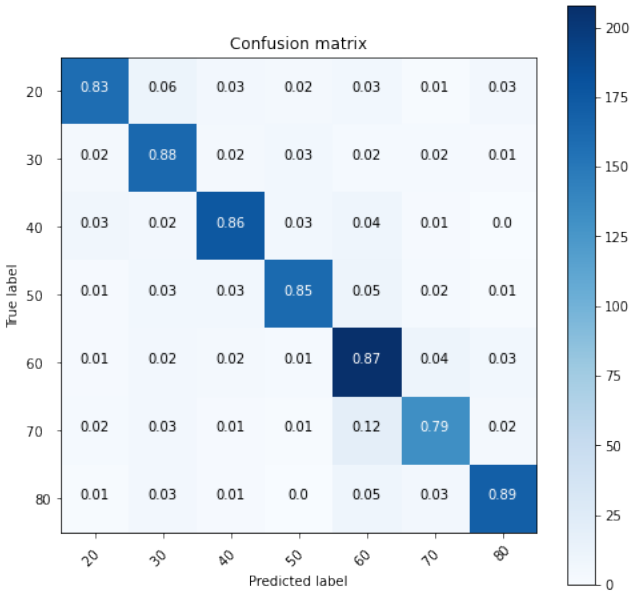

FIGURE 5.12 : Matrice de confusion de la première configuration pour θ

Mesures / Classes	**Précision**	**Rappel**	**F1-mesure**	**Support**
20	0.89	0.83	0.86	191
30	0.81	0.88	0.85	185
40	0.88	0.86	0.87	205
50	0.89	0.85	0.87	189
60	0.78	0.87	0.83	238
70	0.85	0.79	0.82	167
80	0.90	0.89	0.89	192

TABLE 5.10 : Rapport des mesures d'évaluation de la première configuration pour la classification de θ

Deuxième configuration

Hyper-paramètres	Valeurs
Nombre de couche cachée	2
Nombre de neurone pour chaque couche cachée	617
Nombre de neurone pour la couche de sortie	7
Fonction d'activation couches cachées	ReLU
Fonction d'activation couche sortie	Softmax
Algorithme d'optimisation	Adam
Taux d'apprentissage (Learning rate)	ReduceLROnPlateau(0.1 jusqu'à 0.00001)
taille du lot (batch size)	60
Fonction d'erreur	Entropie-catégorielle
Nombre d'itérations (epochs)	350

TABLE 5.11 : Deuxième configuration de réseau implémenté pour la classification de θ

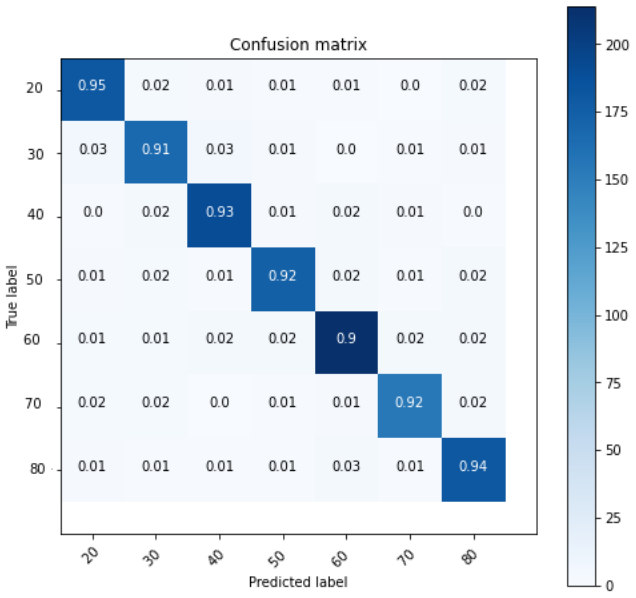

FIGURE 5.13 : Matrice de confusion de la deuxième configuration pour θ

Mesures / Classes	**Précision**	**Rappel**	**F1-mesure**	**Support**
20	0.91	0.95	0.93	191
30	0.91	0.91	0.91	185
40	0.93	0.93	0.93	205
50	0.94	0.92	0.93	189
60	0.93	0.90	0.91	238
70	0.94	0.92	0.93	167
80	0.91	0.94	0.92	192

TABLE 5.12 : Rapport des mesures d'évaluation de la deuxième configuration pour la classification de θ

Distance radiale R

Les résultats de simulation pour la prédiction de la distance radiale R sont représentés respectivement dans les tableaux 5.13 et 5.15 pour les deux confgurations 5.14 et 5.16.

Première configuration

Hyper-paramètres	**Valeurs**
Nombre de couche cachée	2
Nombre de neurone pour chaque couche cachée	617
Nombre de neurone pour la couche de sortie	10
Fonction d'activation couches cachées	ReLU
Fonction d'activation couche sortie	Softmax
Algorithme d'optimisation	Adam
Taux d'apprentissage (Learning rate)	0.1
taille du lot (batch size)	60
Fonction d'erreur	Entropie-catégorielle
Nombre d'itérations (epochs)	350

TABLE 5.13 : Première configuration de réseau implémenté pour la classification de R

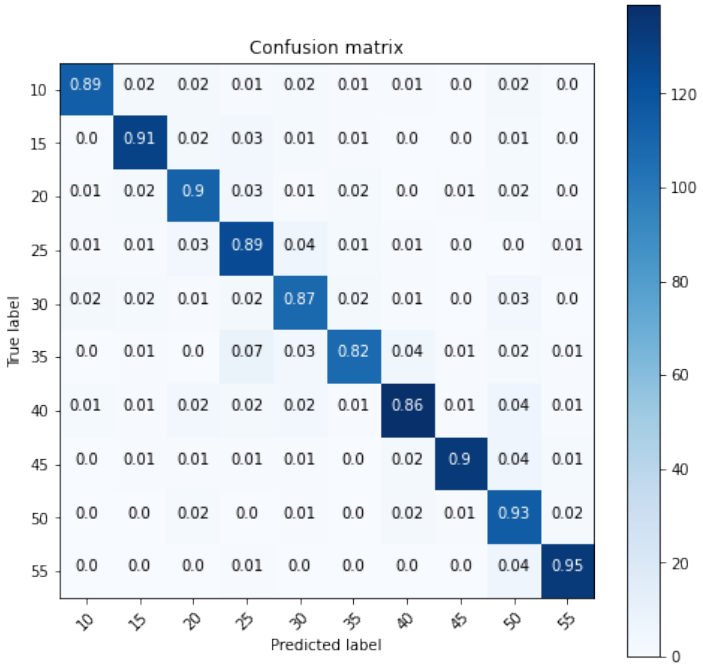

FIGURE 5.14 : Matrice de confusion de la première configuration pour R

Mesures / Classes	**Précision**	**Rappel**	**F1-mesure**	**Support**
10 mm	0.94	0.89	0.92	128
15 mm	0.92	0.91	0.92	141
20 mm	0.86	0.90	0.88	125
25 mm	0.83	0.89	0.86	140
30 mm	0.84	0.87	0.85	126
35 mm	0.92	0.82	0.87	131
40 mm	0.91	0.86	0.88	162
45 mm	0.97	0.90	0.93	149
50 mm	0.79	0.93	0.85	124
55 mm	0.95	0.95	0.95	141

TABLE 5.14 : Rapport des mesures d'évaluation de la première configuration pour la classification de R

Deuxième configuration

Hyper-paramètres	Valeurs
Nombre de couche cachée	2
Nombre de neurone pour chaque couche cachée	617
Nombre de neurone pour la couche de sortie	10
Fonction d'activation couches cachées	ReLU
Fonction d'activation couche sortie	Softmax
Fonction d'optimisation	Adam
Taux d'apprentissage (Learning rate)	ReduceLROnPlateau(0.1 jusqu'à 0.00001)
Taille du lot (batch size)	60
Fonction d'erreur	Entropie-catégorielle
Nombre d'itérations (epochs)	350

TABLE 5.15 : Deuxième configuration de réseau implémenté pour la classification de R

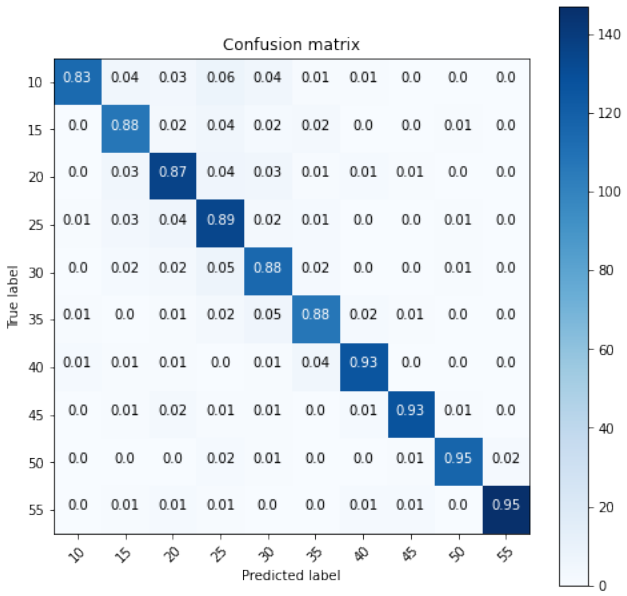

FIGURE 5.15 : Matrice de confusion de la deuxième configuration pour R

Mesures / Classes	**Précision**	**Rappel**	**F1-mesure**	**Support**
10 mm	0.97	0.83	0.89	138
15 mm	0.85	0.88	0.87	121
20 mm	0.86	0.87	0.86	156
25 mm	0.80	0.89	0.84	151
30 mm	0.83	0.88	0.86	130
35 mm	0.88	0.88	0.88	121
40 mm	0.93	0.93	0.93	136
45 mm	0.96	0.93	0.95	136
50 mm	0.97	0.95	0.96	123
55 mm	0.99	0.95	0.97	155

TABLE 5.16 : Rapport des mesures d'évaluation de la deuxième configuration pour la classification de R

5.5.2 Discussion

Le tableau 5.17 regroupe les résultats obtenus des différents configurations pour tous les paramètres. Il contient l'exactitude, le taux d'erreur et le support. les valeurs moyennes de ces métriques sont presque égale :

$$\text{Précision}_{moyenne} = \text{Rappel}_{moyenne} = \text{F-mesure}_{moyenne}$$

ce qui décrit l'homogénéité de notre solution. Ces résultats montrent que la deuxième configuration donne des meilleurs résultats pour tous les paramètres. Cela est dû à :

- L'amélioration du nombre d'epochs (voir la figure 5.16 qui montre que le nombre d'epochs améliore la précision à partie de la valeur 100 pour la deuxième configuration ce qui constitue le meilleur choix dans notre étude),
- Le taux d'apprentissage qui atteint des valeurs minimales
- La fonction d'optimisation (on remarque que l'utilisation de la fonction ReduceLrOnPlateau a permis de réduire le taux d'erreur global et d'augmenter la précision pour les classifieurs).

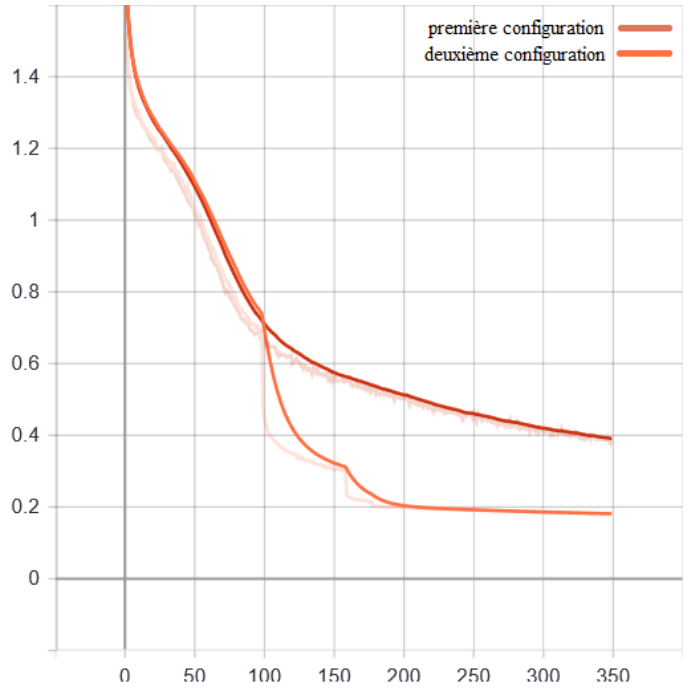

FIGURE 5.16 : Erreur vs. Nombre d'epochs (cas de θ)

L'exactitude dans notre solution avoisine à chaque fois 91% ce qui peut être considère comme un bon résultat de prédiction avec une erreur de 20%. Les différentes matrices de confusion nous renseignent sur la performance de notre solution et notamment par rapport aux vrai positif et faux négatif qui représentent des pourcentages importants. Nous avons remarqué aussi que la prédiction est plus précise dans de la cas R et θ par rapport à α qui est dû sans doute au nombre de classes pour le paramètre α.

	Exactitude	**Taux d'erreur**	**Support**
1^{ere} config classification α	83,75%	53,74%	1367
2^{eme} **config classification** α	92,91%	24,32%	1367
1^{ere} config classification θ	84,28%	40,48%	1367
2^{eme} **config classification** θ	92,02%	17,67%	1367
1^{ere} config classification R	89,03%	34,19%	1367
2^{eme} **config classification** R	91,92%	21,7%	1367

TABLE 5.17 : Résultats des deux configurations pour classification de R

5.6 Conclusion

Ce chapitre a été consacré à la présentation et la discussion des résultats de la simulation. Nous avons remarqué que chaque étape, depuis l'acquisition des données à la prédiction de la position de la tumeur en passant par le traitement des données ainsi que la configuration du classifieur neuronal, a son impact sur ces résultats. Les résultats obtenus représentent particulièrement de bonnes performances qui peuvent encourager a poursuivre l'amélioration de cette solution.

Chapitre 6

Conclusion Générale

Ce mémoire s'inscrit dans l'objectif d'aide au diagnostique du cancer du sein par une méthode inspirée de la thermographie médicale. On s'est concentré dans notre travail sur la dernière étape de conception d'un système de diagnostique médical (SADM), soit la classification des états de positions de tumeurs dans un sein, par les réseaux de neurones artificiels multicouches (RNMC).
Nous avons d'abord commencé par présenter la maladie de cancer du sein et ses techniques de dépistage. Ensuite, nous avons fourni un aperçu sur les propriétés fondamentales de l'apprentissage automatique et de réseau de neurones artificiels que nous avons utilisé pour réaliser notre travail.
Nous avons ensuite tracé une architecture générale du système tout en détaillons ses principaux composants. Nous nous sommes focalisé dans notre travail sur l'analyse des mesures de températures en s'inspirant de la technique de thermographie médicale afin d'étudier ses performances dans le domaine du diagnostique du cancer du sein, car cette technique a l'avantage d'être moins cher, accessible et non-nocive à la santé des malades. Cette solution pourrait présenter aux experts un outil rapide et performant d'aide à la décision finale. La solution proposée démontre de bonne performances, néanmoins, des améliorations sur les données et les algorithmes peuvent êtres apportées pour plus d'efficacité et notamment décisionnelle.
Enfin, ce travail ouvre le champs sur plusieurs travaux de futurs tels que l'amélioration du modèle, la réduction des dimensions des entrées et l'amélioration de la performance.

Bibliographie

[1] *Le larousse médical (édition 2012)*. 2012.

[2] Le langage de programmation python, 2019. Last accessed 8 February 2019.

[3] Pycharm ide, 2019. Last accessed 8 February 2019.

[4] Estimation nationale de l'incidence et de la mortalité par cancer en france entre 1980 et 2012. Étude à partir des registres des cancers du réseau francim. partie 1 : tumeurs solides. Juillet 2013.

[5] Salima Akkouche. cancer du sein : les inquiétantes statistiques des spécialistes. *Algérie 360*, 2019.

[6] David B Aronow, Thomas H Payne, and S Pierre Pincetl. Postdoctoral training in medical informatics : a survey of national library of medicine-supported fellows. *Medical Decision Making*, 11(1) :29–32, 1991.

[7] Haute autorité de santé. Dépistage et prévention du cancer du sein, 2015. Last accessed 8 February 2019.

[8] Eta S Berner, Richard S Maisiak, C Glenn Cobbs, and O David Taunton. Effects of a decision support system on physicians' diagnostic performance. *Journal of the American Medical Informatics Association*, 6(5) :420–427, 1999.

[9] Valentin Bisson. Algorithmes d'apprentissage pour la recommandation. 2013.

[10] D Chen and Douglas Stow. The effect of training strategies on supervised classification at different spatial resolutions. *Photogrammetric Engineering and Remote Sensing*, 68(11) :1155–1162, 2002.

[11] Xavier Dupré. réseaux de neurones artificiels , wikistat. Last accessed 16 September 2019.

[12] JF Ferlay. Globocan 2000. cancer incidence, mortality and prevalence worldwide, version 1.0. *IARC cancerbase*, 2001.

[13] Adrien Guille. Évaluation en classification supervisée. Last accessed 02 July 2019.

[14] Jean-Louis HARTENBERGER. Nous, les mammifères, paris, le pommier. 2013.

[15] Gao Huang, Shiji Song, Jatinder ND Gupta, and Cheng Wu. Semi-supervised and unsupervised extreme learning machines. *IEEE transactions on cybernetics*, 44(12) :2405–2417, 2014.

[16] Mohamed Khireddine Kholladi et al. Vie artificielle, analyse, traitement et fouille de donnees.

[17] Zineddine Kouahla. Plateforme de développement pour l'internet des objets (ido) avec un apprentissage automatique. 2019.

[18] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553) :436–444, 2015.

[19] Mohammad Saeid Mahdavinejad, Mohammadreza Rezvan, Mohammadamin Barekatain, Peyman Adibi, Payam Barnaghi, and Amit P Sheth. Machine learning for internet of things data analysis : A survey. *Digital Communications and Networks*, 4(3) :161–175, 2018.

[20] Jean Tafforeau Marina Puddu. Opportunité de dépistage du cancer du sein chez les femmes de 40 à 49 ans. *IARC cancerbase*, 2005.

[21] Nick McClure. *TensorFlow machine learning cookbook*. Packt Publishing Ltd, 2017.

[22] Moïse Namer, Michel Héry, Daniel Serin, Marc Spielmann, and Joseph Gligorov. *Cancer du sein : Compte rendu du cours supérieur francophone de cancérologie (Saint-Paul-de-V Vence, ence, 18-20 janvier 2007)*. Springer, 2007.

[23] Kazar Okba. Intelligence artificielle et ses applications. Last accessed 02 July 2019.

[24] Wu Tao and Qin Kun. Cloud model method in disaster loss assessment. In *2009 International Forum on Information Technology and Applications*, volume 2, pages 673–676. IEEE, 2009.

[25] Claude Touzet. *les réseaux de neurones artificiels, introduction au connexionnisme*. 1992.

[26] Isabelle V. La thermographie pour dépister le cancer du sein, 2017. Last accessed 12 March 2019.

[27] Tony Yiu. understanding random forest, 2019. Last accessed 16 September 2019.